WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Human Activity Identification Using Pose Estimation

[1]Shubhangi Kedar, [2]Shivkanya Doiphode, [3]Ankita Gaikwad, [4]Divya Kurumkar, [5]Swati Shekapure

[1,2,3,4,5]Marathwada Mitra Mandal's College of Engineering Pune, Maharashtra, India

ABSTRACT

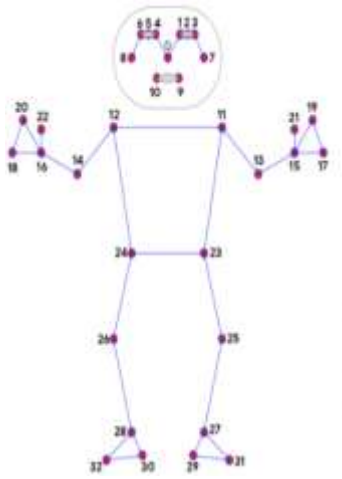
Recognition of human activity is important for interpersonal communication and human-to-human /machine interaction. One of the primary research topics in the fields of computer vision and machine learning is the machine's capacity to recognise other people's actions. It is now possible to gather and retain data on various elements of human movement under the conditions of free living and identify the action performed thanks to the development of Pose estimation together with Classification algorithms, which can be utilized on images/video input. This method could be applied to automated activity profiling systems that generate a continuous record of activity patterns over long periods of time [1]. These methods for activity profiling rely on classification algorithms that can effectively detect various activities using motion data. The posture estimate approach (mediapipe) is reviewed in this study and has been used to categorize typical activities and/or identify falls using body-joint data. The review is organized in accordance with the various analytical methods and exemplifies the range of strategies that have previously been used in this area. Although there has been a lot of development in this crucial field, there is still a lot of room for improvement, particularly in the application of cutting-edge categorization algorithms to issues involving a wide range of activities [2].

Introduction

Human activities may be categorized into three categories based on the inherent hierarchical structure that reflects the different levels of them. Walking, chatting, standing, and sitting are just a few examples of common interior activities that include movements. Because it affects wellbeing, the identification of human activity has many applications [3].They might also tasks that are more narrowly focused, such those done in a kitchen or on a manufacturing floor. Using Pose estimation and a classification algorithm, the challenge of human activity recognition entails determining what a person is doing based on a trace of their movement. It is a difficult problem since there are several motions needed to do certain activities in a broad sense. Studying the identification, interpretation, and analysis of behaviors peculiar to human movement is known as "human activity recognition." This project will examine what the user is doing when viewing the video or image input. Pose estimation and classification algorithms will be used to analyze the data collection and discover human activities.

Proposed System

Pose estimation - Pose Estimation is a computer vision task that infers the pose of a person or object in an image or video. Pose estimation is the problem of determining the position and orientation of a camera relative to a given person or object. A number of keypoints on a specific object or person are typically identified, located, and tracked in order to accomplish this. This could be corners or other notable features on an object. These key points also correspond to significant joints in humans, such as the elbow and knee. Our machine learning models' objective is to identify these keypoints in images [4].



0. nose
1. left_eye_inner
2. left_eye
3. left_eye_outer
4. right_eye_inner
5. right_eye
6. right_eye_outer
7. left_ear
8. right_ear
9. mouth_left
10. mouth_right
11. left_shoulder
12. right_shoulder
13. left_elbow
14. right_elbow
15. left_wrist
16. right_wrist
17. left_pinky
18. right_pinky
19. left_index
20. right_index
21. left_thumb
22. right_thumb
23. left_hip
24. right_hip
25. left_knee
26. right_knee
27. left_ankle
28. right_ankle
29. left_heel
30. right_heel
31. left_foot_index
32. right_foot_index

Fig 1. Body key points

Dataset - Based on the mediapipe analysis we create our CSV file to train our Classification Algorithm and Prediction is made about the human activity. The admin can use webcam for live input or upload video input and according to the details the prediction will be made. Logistic Regression was used to predict the action.

	class	x1	y1	z1	v1
0	Standing	0.441929	0.179151	-0.768026	0.999997
1	Standing	0.441217	0.179151	-0.949387	0.999997
2	Standing	0.440749	0.179168	-0.946997	0.999997
3	Standing	0.440405	0.179345	-0.940684	0.999997
4	Standing	0.440183	0.179372	-0.959358	0.999996

Fig 2. Dataset format

Class - action information/name (6 classes)

x and y - Landmark coordinates normalized to [0.0, 1.0] by the image width and height of the image respectively.

z: Represents the landmark depth with the depth at the midpoint of hips being the origin, and the smaller the value the closer the landmark is to the camera. The magnitude of z uses roughly the same scale as x.

visibility: visibility: number between [0.0, 1.0] that represents the possibility that the landmark will be visible(present and not occluded) in the image.

```
Namaste          468
Standing         424
Hands_On_Waist   420
Bend_Down        331
Left_Hand_Up     272
Right_Hand_Up    252
```

Fig 3. Actions in dataset

System Architecture

1. **Input Data:** The input data is received through camera in form of Image/Video
2. **Data preprocessing:** The aim of preprocessing is to find the most informative set of features to improve the performance of the classifier.
3. **Pose Estimation:** Pose estimation is a computer vision task that infers the pose of a person or object in an image or video.
4. **Feature Extraction :** extract features and classify activities from raw data.

5. **Draw skeleton:** The coordinates of the skeleton joints are used to represent Human Action.

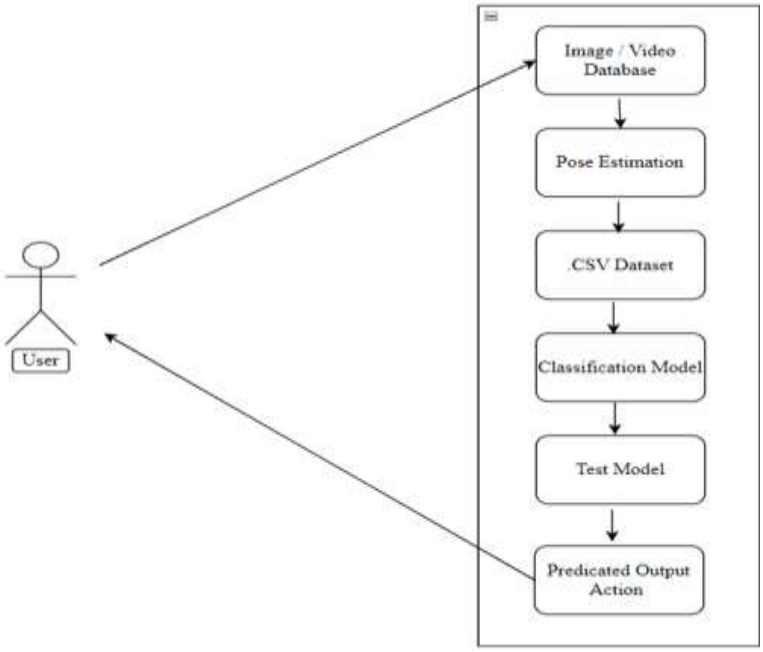


Fig 4. Architecture

Pose Estimation

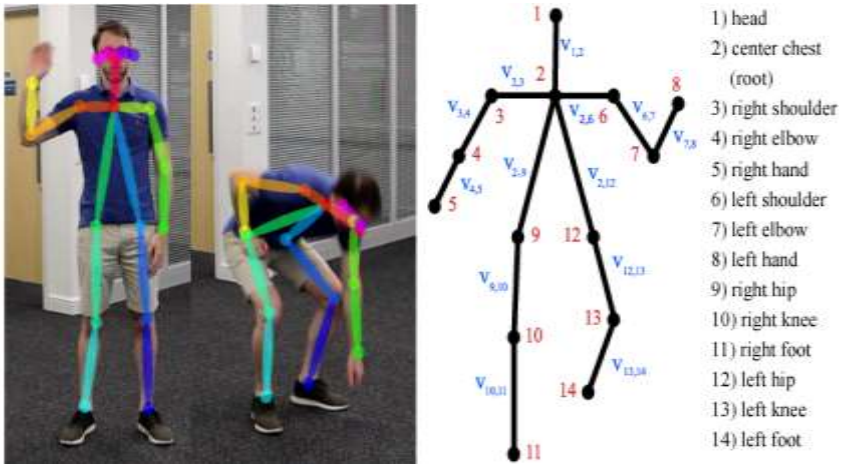


Fig 5. Body Keypoints

Pose estimation is a computer vision task that infers the pose of a person or object in an image or video. Pose estimation is the problem of determining the position and orientation of a camera relative to a given person or object. Pose estimation is a computer vision technique to track the movements of a person or an object. Finding the locations of important points for the given objects is typically how this is done. We can compare different motions and postures based on these fundamentals and draw conclusions [5].

Types of Pose Estimation:

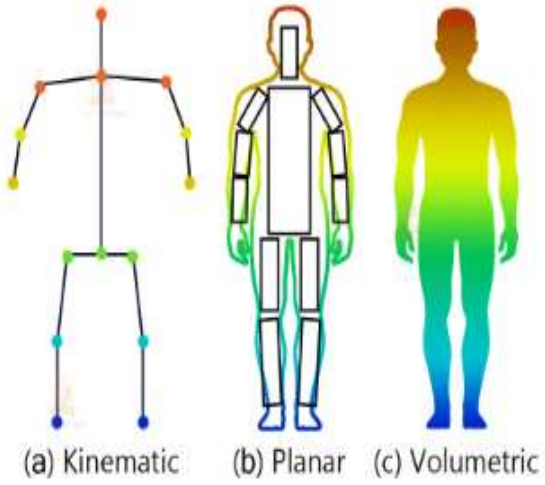


Fig 6. Pose Estimation Type

Kinematic: It is also called the Skeleton model, this representative includes a set of key points (joints) like ankles, knees, shoulders, elbows. In this project we have used this this type of pose estimation.

Planar: it represents the appearance and shape of a human body, where body parts are displayed with boundaries and rectangles of a person's contour.

Volumetric: It consists of various geometric shapes, such as cylinders, conics, and triangulated meshes.

Preprocessing

Skeleton detection is a crucial technology for apps that include image as part of the scene. It is need of the time to keep track of as many people that appear in the image. Now, with image processing combined with AI generates a dots and lines joining those dots. Image classification accepts an input as a picture with any extension and processes and manipulate it for specific classification. AI or Vision sees input as picture to exhibit of pixels and it relies upon the picture goals [5].

Feature Extraction

Skeleton shaped object is essential for extracting features and the properties from the segment object. One of the aims of using the system proposed is to specify different parts and its position for human body. Each part is separately analyzed for feature extraction. The method divides the body into different blocks each one consist of different pose known by the system .

Classification Algorithm

We define the activity classification problem as a multiple class classification problem, which can be modelled using different machine learning classification and regression algorithms. The classification algorithm takes 33 body keypoints dataset (x-axis, y-axis z-axis and visibility values of each point) as input for the model's training and testing. We used a supervised learning approach as we have labelled dataset containing body keypoints with an activity label. Among all the algorithms, we use logistic regression which provides 100% accuracy.

Logistic Regression - It is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. In simple words, the dependent variable is binary in nature having data coded as either 1 (stands for success/yes) or 0 (stands for failure/no). Mathematically, a logistic regression model predicts P(Y=1) as a function of X. Logistic regression models the data using the sigmoid function [3].

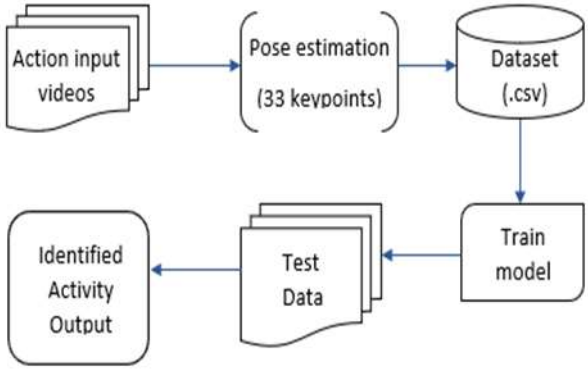


Fig 7. System Workflow

Components

Mediapipe- Mediapipe is a cross-platform library developed by Google that provides amazing ready-to-use ML solutions for computer vision tasks. OpenCV library in python is a computer vision library that is widely used for image analysis, image processing, detection, recognition, etc.MediaPipe Pose is a single-person pose estimation framework. It uses BlazePose 33 landmark topology. BlazePose is a superset of COCO keypoints, Blaze Palm, and Blaze Face topology. It works in two stages – detection and tracking. As detection is not performed in each frame, MediaPipe is able to perform inference faster**.** There are three models in MediaPipe for pose estimation.

Opencv- In OpenCV, images are converted into multi-dimensional arrays, which greatly simplifies their manipulation. For instance, a grayscale image is interpreted as a 2D array with pixels varying from 0 to 255. OpenCV is the huge open-source library for the computer vision, machine learning, and image processing and now it plays a major role in real-time operation which is very important in today's systems. By using it, one can process images and videos to identify objects, faces, or even handwriting of a human.OpenCV has a bunch of pre-trained classifiers that can be used to identify objects such as hand, legs faces, eyes, etc

Evaluation Metrics

Precision Precision(P) – It is the ratio of the number of true positives (Tp) to the sum of false positives (Fp) and true positives. It can also be defined as how many images classified into this class belong to this class.

Recall (R) - It is the ratio of the number of true positives (Tp) to the sum of false negatives (Fn) and true positives. It can also be defined as how many images that belong to this class are classified into this class.

F1-Score – It is calculated as the harmonic mean of recall and precision. Equation calculates it.

Confusion Matrix - It is a two-dimensional matrix used to measure the overall performance of the machine learning classification algorithm. In the matrix, each row is associated with the predicted activity class, and each column is associated with the actual activity class. The matrix compares the target activity with the activity predicted by the model. This gives a better idea of what types of errors our classifier has made [3].

Results and Discussion

We have used Logistic regression, ridge classifier, random forest classifier, gradient boosting classifier and got 100% accuracy for Logistic regression, random forest classifier, gradient boosting classifier and 99% for ridge classifier. We used logistic regression as our final model for prediction.

Algorithm	Precision	Recall	F1-Score	Support
Logistic regression	1.0	1.0	1.0	651
Ridge classifier	0.99	0.99	0.99	651
Random forest classifier	1.0	1.0	1.0	651
Gradient boosting classifier	1.0	1.0	1.0	651

Fig 8. Evaluation of algorithms

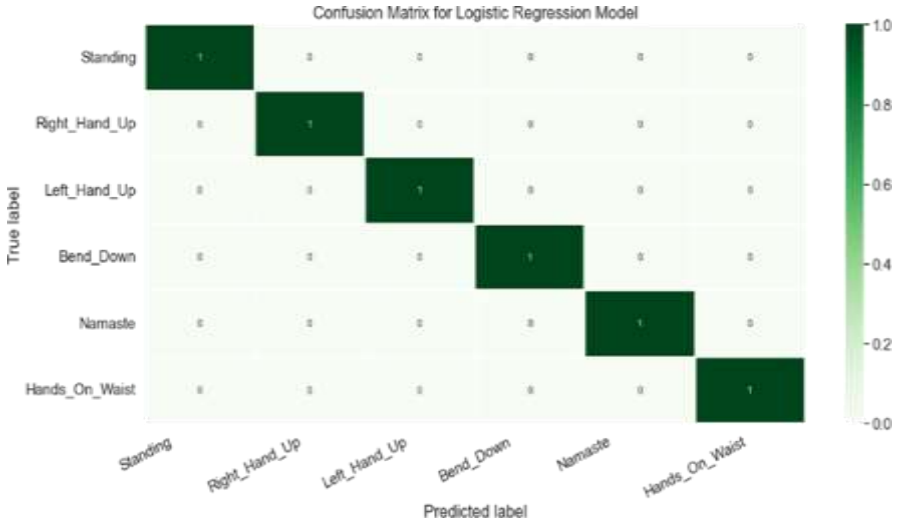


Fig 9. Confusion Matrix for Logistic Regression

Fig 10. Action identified as Namaste

Conclusion

In this project we have introduced Human Activity Identification technique which will detect the activity being performed in the input video. We have proposed a Human Activity Identification system based on pose estimation and Classification algorithm. This system will combine the results of the 3D pose estimation model with the classification technique for better and more accurate detailed action prediction. We have used Logistic regression, ridge classifier, random forest classifier, gradient boosting classifier and got 100% accuracy for Logistic regression, random forest classifier, gradient boosting classifier and 99% for ridge classifier. We used logistic regression as our final model for prediction.

References

[1] Palwe, A., Shiravale, S. and Shekapure, S., 2022. Image Captioning using Efficient Net. Journal of Optoelectronics Laser, 41(7), pp.1259-1270

[2] Shekapure, Swati, and Dipti D. Patil. "Enhanced e-Learning experience using case based reasoning methodology." International Journal of Advanced Computer Science and Applications 10, no. 4 (2019)

[3] B. M. V. Guerra, S. Ramat, R. Gandolfi, G. Beltrami and M. Schmid, "Skeleton data preprocessing for human pose recognition using Neural Network*," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 2020, pp. 4265-4268, doi: 10.1109/EMBC44109.2020.9175588

[4] J A. Gupta, A. Kembhavi and L. S. Davis, "Observing Human Object Interactions: Using Spatial and Functional Compatibility for Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 10, pp. 1775-1789, Oct. 2009, doi: 10.1109 /TPAMI.2009.83.

[5] J. Shotton et al., "Realtime human pose recognition in parts from single depth images," CVPR 2011, Providence, RI, 2011, pp. 1297-1304, doi: 10.1109/CVPR.2011.5995316.

[6] R. Liu, T. Chen and L. Huang, "Research on human activity recognition based on active learning," 2010 International Conference on Machine Learning and Cybernetics, Qingdao, 2010, pp. 285-290, doi: 10.1109/ICMLC.2010.5581050.

[7] E. Bulbul, A. Cetin and I. A. Dogru, "Human Activity Recognition Using Smartphones," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technolo-gies (ISMSIT), Ankara, 2018, pp. 1-6, doi: 10.1109/ISMSIT.2018.8567275

[8] Yang Wang, Hao Jiang, M. S. Drew, ZeNian Li and G. Mori, "Unsupervised Discovery of Action Classes," 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 2006, pp. 1654-1661, doi: 10.1109/CVPR.2006.321.

[9] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensorbased activity recognition: A survey,” Pattern Recognit. Lett., vol. 119, pp. 3–11, Mar. 2019

[10] C. Xu, D. Chai, J. He, X. Zhang, and S. Duan, “InnoHAR: A deep neural network for complex human activity recognition,” IEEE Access, vol. 7, pp. 9893–9902, 2019.

[11] T. Zebin, P. J. Scully, N. Peek, A. J. Casson, and K. B. Ozanyan, “Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition,” IEEE Access, vol. 7, pp. 133509–133520, 2019.