Body Posture Detection Using Computer Vision

Anand Upadhyay, Ketan Chaudhari, Pradip Bhere, Jerin Thomas

Department of Computer Engineering, Ramrao Adik Institute of Technology. Nerul, India.

ABSTRACT

Vision based human posture recognition has fascinated many researchers due to its critical challenges and a variety of applications. The applications range from simple posture recognition to complicated behaviour understanding in surveillance systems. This leads to major development in the techniques related to human body posture representation and recognition. This paper discusses applications, general framework of human body posture recognition. In this paper we present a approach for Body Posture detection in OpenCV Using OpenPose MobileNet Technology. The paper also emphasizes its advantages and disadvantages. The human body posture recognition domain has been active for more than two decades, and has provided a large amount of literature. The paper also includes some other methods of human posture detection with its limitations and advantages.

Keywords: *Body Posture Detection, OpenCV.*

I. INTRODUCTION

Computer Vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images and videos. Therefore, Human Body posture detection using computer vision deals with the analysis and high-level understanding of the body posture, detection of the presence of human body in a surrounding and the detailed analysis of the its movements with respect to the surrounding and the various human movements from digital images and videos.The motivation for further developments in this field comes from the desire to improve human computer interaction and due to the promising applications of this field.

Vision based human body posture recognition is a systematic approach to understand and analyse the movement of people in camera captured content. It also comprises fields such as Biomechanics, Machine Vision, Image Processing, Artificial Intelligence and Pattern Recognition. It is an interdisciplinary challenging field having grand applications with social, commercial, and educational benefits. A wide spectrum of applications demands human motion recognition.

The applications are spread over domains like sports, medical, surveillance, content based video storage and retrieval, man-machine interfaces, video conferencing, art and entertainment, and robotics. Some of the applications for highlighting the potential impact of human motion recognition are discussed here.

- ***Smart Surveillance:*** In today's surveillance systems, video contents are viewed continuously by human operators. With the increasing number of cameras, it is impossible for humans to monitor all the contents 24 X 7. Generally, the contents are viewed after a mishap to analyse the event. So, there is an intense requirement of smart surveillance hi ok systems from the security agencies. Smart surveillance systems can analyse an event online and provide appropriate intimation using computer based human motion and behavioural analysis. Smart surveillance is required for access control in special areas like military territory, distant human identification, counting the persons and congestion analysis, detection of abnormal behaviour at shopping malls and Railway Stations.
- ***Behavioural Biometrics:*** Nowadays, the use of the gait pattern as a biometric has become popular. The main reason is that the recognition of the gait pattern does not require subject cooperation as compared to the other biometrics.
- ***Gesture and Posture Recognition and Analysis:*** For a more advanced natural interface with computers and computerized systems, human gesture and posture recognition is an important key. It has promising applications such as gaming, sign language recognition and controlling devices.
- ***Robotics:*** Human motion analysis plays an important role in robotics for humanoid robot control, to imitate human motions in a robot in virtual and augmented environments.
- ***Medical:*** The medical field uses human motion recognition for the study and analysis of

Orthopaedics, Neurology, Musculoskeletal disorders, body posture, and fitness. It is also useful to design intelligent systems to assist elderly people and physically / mentally disabled ones.

• ***Sports and Exercise:*** In sports, motion recognition is useful to analyze athletic movements and to design affordable and efficient frameworks for training. An environment for rehabilitation exercise with a feedback system at remote places or in the presence of an expert is designed. Dao proposed a monitoring system for the exercises of elderly people. These kinds of systems will definitely be useful for patients and old age people.

• ***Art and Entertainment:*** Motion recognition is useful in analyzing, learning, and an emotional understanding of artistic dance movements as in dances like Bharatnatyam, and Salsa. Kale and Patil have recognized Bharatnatyam dance sequence from depth data. This also helps to increase the effectiveness of a scene, and the alteration of movements required for quality and the impact of acting.

II. PROPOSED WORK

Ideal posture indicates proper alignment of the body's segments such that least amount of energy is required to maintain a desired position. Good posture optimizes breathing and improves the circulation of bodily fluids. With a simple posture monitoring system as described in this paper, one can monitor the posture easily and thereby correct it in the long run.

Few of the methods used for Human Posture detection are:

A. Bottom up approach with Discrete HMM (Hidden Markov Model): In this approach, time-sequential images expressing human action are transformed to an image feature vector sequence by extracting a feature vector from each image. In this algorithm, a mesh feature vector sequence extracted from time-sequential images is converted to a sequence of symbols which correspond to code words in the codebook created by vector quantization. Experimentally, the algorithm wasted on real time-sequential images of tennis actions. The categories under consideration were: forehand stroke, backhand stroke, forehand volley, backhand volley, smash and service.

Limitations:

I. Recognition rate depends on the training pattern i.e. .recognition performance worsens as the number of training patterns decreases. This is due to the fact that HMM parameters mainly depend on the selection of training patterns.

II. A new model needs to be designed for each considered action.

III. Experimentally validated only on 2-D images.

B. Dynamic Time Warping Method: Dynamic time warping (DTW) is one of the algorithms for measuring similarity between two temporal sequences, which may vary in speed. For instance, similarities in walking could be detected using DTW, even if one person was walking faster than the other, or if there were accelerations and decelerations during the course of an observation. It compares two time series and computes the distance between them, even if the two series are shifted in time axis. One of the experimental approaches used the detected position and the configuration of the human body and tracked the relative motion of the body parts relative to the parent body. One dimensional time series for every body part was extracted using the human body decomposition and these time series were compared with the saved templates.

Limitations:

I. This method depends upon the assumption that the body under consideration walks parallel to the image plane.

II. Bodies not parallel to the image plane are not considered.

C. Context free grammar checking: This method describes the temporal and spatial structure of human activities that the system aims to recognize using a context-free grammar (CFG) based representation scheme. A human activity is represented by decomposing it into multiple sub-events and by specifying their temporal, spatial and logical relationships. A sub-event of one activity may be further composed of multiple sub-events of itself, capturing the hierarchical structure of human activities. Once the activities are hierarchically represented, this method is able to recognize them by performing semantic matching between the representation and the observed images from a given sequence in the video. This method was experimentally tested for eight actions – approach, hug, punch, kick and push. Experiments showed that

the system understands continued and recursive composite actions and interactions with a very high recognition rate.

Limitation:

I. Recognition rate for recursive interaction is still less.

D. Human Movement Detection and Identification Using Pyroelectric Infrared Sensors: In this method, human movement is detected and identified using PIR (Pyroelectric Infrared Sensors) based modules having two pairs of orthogonally- aligned PIR sensors. A data collection module consisting of two pairs of PIR sensors whose dual sensing elements are orthogonally aligned a data logger, op-amp circuits and a rechargeable battery. In this method, three PIR-based modules are placed in a hallway, out of which one PIR-based module is placed on the ceiling and two PIR-based modules are placed on opposite walls facing each other.

Limitations:

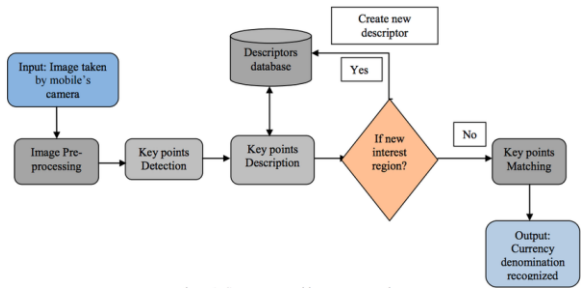
I. It does not operate at temperatures greater than 35 degree Celsius

II. Since PIR sensors sense heat signatures in the room, they may not be effective if the room itself is warm.

III. PIR sensors may turn off if there is very little movement in occupied floors.

IV. All these methods require a number of sensors and hence they are costly.

Our method for posture detection uses a camera and lines of code in Opencv Using Open Pose MobileNet Technology. Which is affordable and easily available components .Below is the flowchart of the technique used by us.

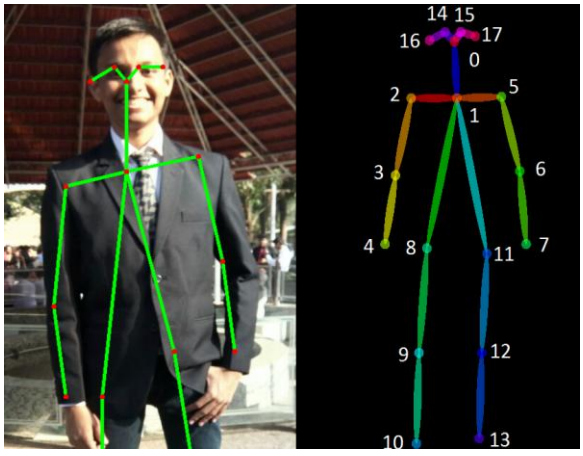


III. METHODOLOGY

The Technique used by our paper is Human Body Posture Detection in OpenCv Using OpenPose Mobile Net Technology. OpenCV is a library of programming functions aimed for real-time computer vision. It is cross-platform and free. It supports deep learning frameworks such as Tensor Flow. The basic version of OpenCV was released to public IEEE Conference on Computer Vision and Pattern Recognition. It was built in C++ but has bindings in Python and MATLAB.

MobileNets are small, low-latency, low powered models which are used to meet the resource constraints of a variety of use cases. It comes under the family of mobile-first vision computer models designed to effectively maximize accuracy while being mindful of the resources for an on device or embedded application. It is a class of efficient modules for mobile and embedded vision applications. They are based on a streamlined architecture that uses depth-wise separable convolutions to build light weight Deep-Neural Networks. Effectiveness of MobileNets can be demonstrated across a wide range of applications and use cases including object detection, Finegrain Classification, Face attributes, Landmark Recognition and large scale geo-localization. It can be run on mobile devices with help of Tensor Flow Mobile, it enables on-device machine learning inference with low latency. With help of Android Neural Networks API. It focuses on optimizing for latency but also yields small networks.

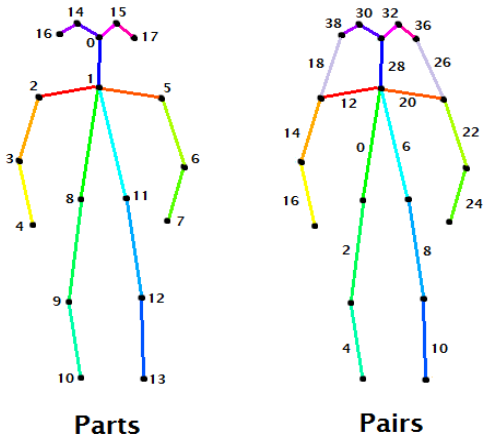
In the image below we detected the position and orientation of an object by detecting the key point locations that describe the object. We focused on human body posture estimation where it is required to detect and localize the major parts or joints of the body (e.g. Shoulders, ankle, knee, wrist etc.). We can implement the idea using webcam as well as pre-available images. It gives more accuracy in second case.



Sample Output A (Standing)

Sample Output B (Sitting)

Below is the image of the key points in the body and the corresponding pairs



MobileNet Architecture:

Firstly we will describe the core layers that MobileNet is built on which are depth wise separable filters. For MobileNets the depthwise Convolution applies a single filter to each input channel. The pointwise convolution then applies a 1×1 convolution to combine the outputs the depthwise convolution. A standard convolution both filters and combines inputs into a new set of outputs in one step. The depth wise separable convolution splits this into two layers, a separate layer for filtering and separate layer for combining. MobileNet model is based on depth wise separable convolutions. It factorizes a standard convolution into a depthwise convolution. A 1x1 convolution is called pointwise convolution. In depthwise convolution applies a single filter to each input channel. The pointwise convolution then applies a 1x1 convolution to combine the outputs of the depthwise convolution. Depthwise separable convolutions are made of two layers: depthwise convolutions and pointwise convolutions. The combination of depthwise convolution and 1x1(pointwise) convolution is called depthwise separable convolution. Depthwise convolution with one filter input channel can be written as:

$G_{k,l,n}= \Sigma_{i,j,m} K_{i,j,m,n} .F_{k+i-1,l+j-1,m}$

G and F denotes output and input feature map respectively. Here K stands for Kernel. Computational cost of depthwise separable convolution:

$$Dk * Dk * M * Df * Df + M * N * Df * Df$$

Which is the sum of depthwise and 1x1 pointwise convolution. Where Dk the spatial dimension of the kernel K and M is the number of input channels.

It is extremely efficient relative to standard convolution. But, it does not combine input channels to create new features.

MobileNet uses 3x3 depth wise separable convolutions which uses 8 to 9 times less computation than standard convolutions at only a small reduction in accuracy.

Advantages of our proposed system:

1) Human posture detection using OpenCV is cheaper compared to other methods like using pyroelectric infrared sensors.

2) Human posture detection using OpenCV does not require any external attachments on the body or external materials for its working. It only requires a computer system with required software's, webcam and the presence of the object u see consideration.

3) It is more appropriate when it comes to daily practical application as compared to other methods.

4) Requires minimum number of materials as compared to other methods

5) It is more effective in practical applications as compared.

Disadvantages of our proposed system:

1) As compared to the other approach of body posture detection in which we use sensors our system is not travelling friendly or that much portable.

2) In our approach we have to analyse the detected positions of key points which are not as accurate as the systems that use accelerometers and gyroscopes.

3) If there is an obstacle between object and the camera then it will be trouble to mark the key points and detect the posture.

IV. ACKNOWLEDGMENT

We would like to thank the CSI-RAIT Committee, Computer Engineering Department and CSI Research Wing for their constant support.

V. CONCLUSION

Human Body posture recognition is assumed to be the fundamental step of understanding human behavior and this paper gives a brief of detection through computer vision. The recognition of human body posture from video and image content is a significant area of research in computer vision, and has shown considerable progress in the domain. The discussed approach proved to be reliable and robust. This paper describes the human posture detection in OpenCV using OpenPose MobileNet Technology. Also the application of such recognition systems are explained including surveillance, entertainment and healthcare. Still we have to face many issues which are to be resolved to get a perfect model. The advancements achieved should be consolidated in terms of real time conditions and performance to have a firm ground for further research.

VI. REFERENCES

[1] "Deep Learning based Human Pose Estimation using OpenCV ...." 29 May. 2018, https://www.learnopencv.com/deep-learning-based-human-pose-estimation-using-opencv-cpp-python/

[2] "Semantic Representation and Recognition of Continued and ...." 12 Nov. 2008, https://link.springer.com/article/10.1007/s11263-008-0181-1

[3] "Human Movement Detection and Identification Using ... - NCBI." 5 May. 2014, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4063065/

[4] "(PDF) Recognizing Human Action in Time-Sequential ...." https://www.researchgate.net/publication/237076990_Recognizing_Human_Act_ion_in_Time Sequent_ial_Images_using_Hidden_Markov_Mode

[5] "An intuitive guide to Convolutional Neural Networks." 24 Apr.2018, https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050

[6] "An intuitive guide to Convolutional Neural Networks." 24 Apr.2018, https://www.freecodecamp.org/news/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050/