IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Pose and Emotion Estimation based on Activity Recognition

[1]Shahistha Sultana. K, [2]Dr. Suresh L

[1]PG Scholar, [2]Principal
[1]Department of Computer Science & Engineering,
[1]Cambridge Institute of Technology, Bangalore, India

Abstract: Nowadays social Interaction between people tells us about the emotional and mental state the person is currently in. Humans tend to exhibit the emotional and mental state through the body and facial expressions. To predict the emotion of a person, the facial expression is majorly considered while body movements are majorly ignored. In most of the cases, body movements/expressions also reveal the emotional and mental state the person is currently in. In Numerous researches/experiments, only facial expressions are considered to predict the emotional state the person is currently in. Humans tend to fake facial expression, while it is difficult to fake the body expression. This proposed project mainly focuses on the body movements to analyze the emotional state rather than concentrating on facial recognition. To recognition this the human activity is identified from the UCI HAR Smartphone dataset. This data has data that was recorded from 30 volunteers with the age difference of 19-48 years and with 561 measurement variables. The human activity thus identified is compared with the 2-D image and the activity is matched with the 2-D image. This process is done to make sure that the activity matches the activity that has been identified from the UCI HAR smartphone dataset. Each activity has many possible positions, these positions/postures related to the activity that has been identified in the smartphone dataset help us in identifying the emotions that a human could possess while doing the activity.

***Index Terms* - Human Activity Recognition, Pose Estimation, Emotion Estimation based on Pose.**

I. Introduction

The smartphone that is used nowadays is very advanced. They have become our assistants for most of our task we perform during our daily routines. It has become so advanced that they record our heartbeat and they can even play any of our favorite songs as per our wish. These smartphones contain a different type of sensors built-in them. These sensors vary from an accelerometer, Magnetometer, Air Humidity Sensor, GPS, Proximity Sensor, gyroscope, Microphone, Fingerprint Sensor, Pedometer, Touch screen Sensors, Barcode/QR Code sensors, Barometer, Ambient Light Sensor, Heart Rate Sensor, Thermometer, etc. Each sensor performs different types of functions.

An accelerometer is used to detect acceleration, tilt, and vibrations used in determining the movements and orientation along the three axes. App's present in the smartphone makes use of this accelerometer sensor to determine whether the smartphone is in a landscape orientation or if it is in portrait orientation. A gyroscope provides us with the details of the orientation and about the directions like left/right and up/down with great precision. These data collected from the gyroscope helps us in detecting the device's tilt. The gyroscope measures rotation of the device, while the accelerometer cannot measure it. The total orientation is measured by combining the data collected from these sensors like the accelerometer, gyroscope, and compass. Activities performed by a human is recognized with the help of data that is collected from the smartphone's sensors like gyroscope and accelerometer.

The UCI HAR smartphone's dataset consists of data that is recorded by 30 different volunteers who have the smartphone attached to their waist. The age of these subjects varies from19-48 and with 561 measurements variable. After the activity has been identified images related to the activity are loaded and emotions are classified for the respective activity.

II. Literature Survey

The existing system describes how human activity can be predicted using smartphone data [1] and how an emotion can be predicted using video frames.

2.1 Human Activity Recognition using Machine Language Algorithm

Recognition of the human activity is the problem of classifying accelerometer data sequences recorded by specialized accelerometer fitted devices or smartphones into defined known movements [1]. Given numerous measurements every second, the spatial complexity of the results, and the lack of a clear method to link accelerometer data to recognized movements [5], this is a task. Classical solutions to the problem include hand-crafted interfaces from time-series data based on fixed-size windows and testing machine learning models, such as for ensembles of decision trees. Data Mining based techniques are used to exaggerate performance metrics of the human activity recognition system using data preprocessing, feature selection, feature extraction, five different classification types: Decision tree, KNN, Naive Bayes, Support Vector Machine, and Random Forest, used to classify human activities[2][4].

2.2 Emotion recognition using face visible in images/videos

In the existing system, human emotion is identified using photographs of the human face or the video of the person where the facial features are visible clearly [12]. An embedded smartphone camera or a surveillance camera captures an image of the face. Some representative frames are removed from the video, and the facial regions in the images are recovered using a face detection system. To predict the appropriate emotion, on the extracted function are applied classification algorithms for machine learning. [14]

2.3 Emotion recognition using Poses/gestures in images/videos where the face is not visible

Recently, increased attention has been paid to recognizing Spatio-temporal points of interest (STIPs) [3], which has become a key computer vision strategy and topic of study. Local Image Features or Interest Points include a lightweight and abstract representation of patterns in frames [6][9]. In the existing system, images or gestures of the body/pose are used to identify emotions.

III. DESIGN AND ALGORITHM

3.1 Software Requirements

Programming Language: Python
IDE: SPYDER

3.2 Machine learning

Machine learning is a sub-field for artificial intelligence (AI). Machine learning is aimed at understanding the nature of the data and fitting the data into models that people can understand and use easily. While machine learning is a field of computational science, it is different from conventional computing approaches. Throughout traditional computing, algorithms are a set of directly designed instructions used by computers to calculate or overcome problems. Computers can train on data inputs rather than machine learning algorithms and use statistical analysis to produce values within a limited range. Therefore, machine learning allows it easier for computers to build models from sample data to simplify data-based decision-making processes. Every technology customer today has enjoyed learning the machine. Face recognition technology allows social media sites to help users identify and share photos from friends.

Optical character recognition technology (OCR) transforms text images to phone format. Machine learning-driven recommendation engines suggest what movies or TV shows to watch the next users' preferences. Machine learning-driven recommendation engines suggest what movies or TV shows to watch next based on user preferences. Consumers will soon be able to access self-driving cars which rely on learning to navigate my computer. Machine-learning is an emerging area. Common machine learning methods are classified into supervised and unsupervised learning and popular algorithmic approaches to machine learning include neighboring k-nearest algorithms, learning decision making, deep learning random forests, etc.

3.3 Deep Learning

Deep learning is one of the sub-fields of machine learning that is designed based on the brain's structure and neurons called the artificial neural networks (ANN). If we take a large amount of data and continue to train the machine, the performance of the machine improves overtime and Deep learning algorithms tend to perform better compared to other machine learning algorithms.

Deep learning tries to mimic how the human brain can transform into the stimuli of vision and hearing light and sound. A deep learning architecture, inspired by biological neural networks, consists of several layers in an artificial neural network composed of hardware and GPUs.

Deep learning requires the use of a cascade of nonlinear layers of processing units to extract or transform information (or representations) characteristics. The performance of one layer is the following layer results. It is possible to control and use algorithms to classify in-depth learning knowledge or to un-monitor and perform pattern analysis.

Deep learning consumes more data and has managed to beat humans among the machine learning algorithms currently being used and built-in some cognitive tasks. Because of these features, deep learning has become the approach with significant potential in the artificial intelligence domain

3.4 Algorithm Overview

An algorithmic overview of the proposed system is given here:

i. The data from the smartphone dataset is and passed to the LSTM algorithm and the model is trained to identify the activity correctly based on the dataset.
ii. After the output like Standing, sitting, walking, walking upstairs, walking downstairs, and laying of the activity is predicted correctly.
iii. The predicted activity like (only standing and sitting is considered in this paper) is taken as input to the open pose estimate algorithm
iv. In the open pose estimate algorithm, the image is passed through the baseline CNN network first where the possible poses of standing and sitting are considered
v. Then the part confidence map and part affinity map are generated
vi. A greedy bipartite matching algorithm processes the confidence map and the part affinity map to obtain the pose for each image.
vii. Once the pose is recognized (Confirmation step), then the distance between the key points in each pose which include the points on the hand, neck face, and legs, are used to calculate the emotions (only happy & sad emotions are considered here).

IV. SYSTEM ARCHITECTURE

4.1 Activity recognition using LSTM_RNN Algorithm

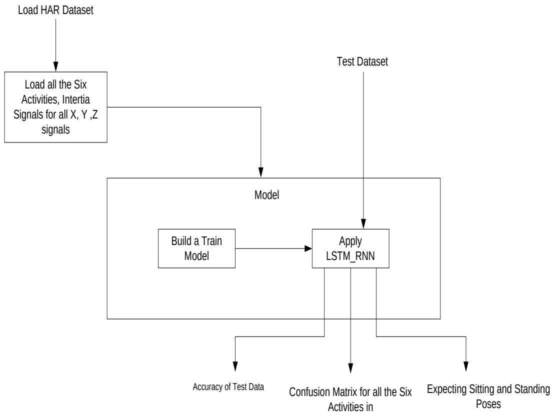

Fig.1: Activity recognition Architecture

Accelerometer data are loaded from the UCI HAR data set and that data is representation time-sliced manner by conversion and reformation. The accelerometer data is then visualized. The multi-dimensional tabular data is reshaped so that the data is accepted. The data is then split up into training data set, validation data set, and test data sets. A Deep Neural network (LSTM_RNN) is defined and then processed by the ML. The machine is trained with this deep neural network for human activity recognition.

4.2 Pose and Emotion estimation using Open pose key point Algorithm

A Human Pose Skeleton is a graphical format representing a person's orientation. Essentially, it's a set of coordinates that can be linked to describe the person's pose. Each skeleton coordinate is known as a part (or joint, or key point). valid relation between two parts is called a pair (or a limb). It Is to be noted that not all combinations of parts give rise to valid pairs.

Load the Keras model for open poses. The Open Pose network extracts features from an image using the first few layers. The characteristics are then fed into two parallel divisions of convolution layers. A set of 18 confidence maps is predicted in the first branch each of which represents a specific part of the human pose skeleton. A 38 Part Fields of Affinity (PAFs) sets are predicted in the second branch, representing the degree of association between parts.

The confidence maps which represent the specific part of the human pose are called as key points and the affinity between these key points becomes the basis for estimating the pose on which the human body's emotions are based.

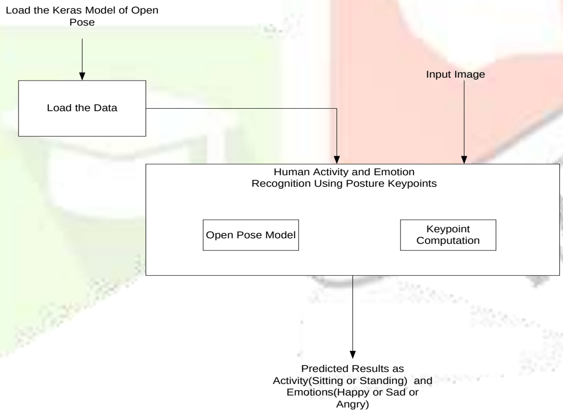

Fig.2: Pose and Emotion Classification Model architecture

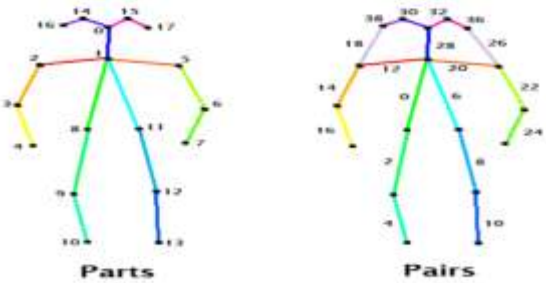

Fig.3: Key Points Skeleton Diagram

V. EXPERIMENT RESULTS

The experiment was able to find the activity clearly and those activities were mapped to the similar kind of images that were given as input to the Open pose key point recognition model.

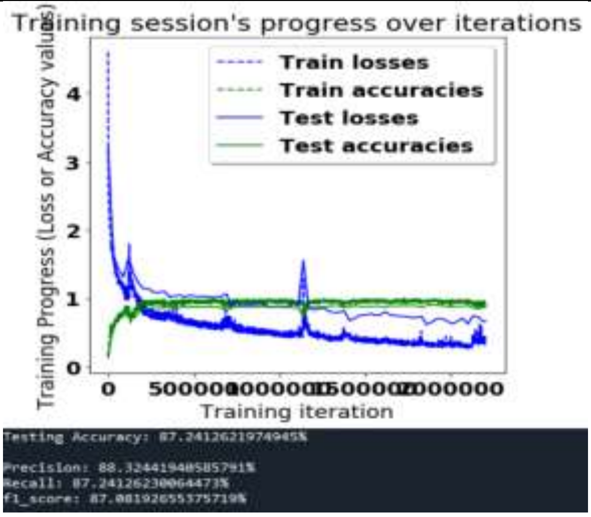

Fig.4: Accuracy and precision graph

Figure 4 gives the accuracy and precision graph for the LSTM algorithm used in finding the human activity from the smartphone dataset

The poses we classified according to the activity and the emotions corresponding to the activity (only happy and sad emotions were identified for the standing and sitting activity). The part confidence map and part affinity map are generated for each image is obtained for each pose.

The pose is recognized (Confirmation step), then the distance between the key points in each pose which include the points on the hand, neck face, and legs, calculates the emotions (only happy & sad emotions are considered here).

Figure 5 describes the happy and sad emotions identified during the experiment while performing the standing activity whereas the figure 6 describes the happy and sad emotions identifies for the sitting activity.

Fig.5: Happy and Sad Emotions for Standing Activity


Fig.6: Happy and sad Emotions for Sitting Activity

VI. Conclusion

This paper presents a way to recognize the activity and makes assumptions that a single activity recognized will have multiple kinds of poses belonging to the same class of the activity recognized. An attempt is made to study all kinds of poses possible for each activity class. Those poses are again evaluated in the poses estimation algorithm and the activity is confirmed again to make sure it matches the activity that is recognized from the UCI HAR dataset. The main idea is to show that each activity won't have only one pose, but many possible poses and humans are prone to exhibit different kinds of emotions related to those activities and not only one emotion.

REFERENCES

[1] Erhan Bulbul; Aydin Cetin; Ibrahim Alper Dogru," Human Activity Recognition Using Smartphones " IEEE 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Dec 2018

[2] Murat Cihan Sorkun; Ahmet Emre Danişman; Özlem Durmaz İncel "Human activity recognition with mobile phone sensors: Impact of sensors and window size" 2018 26th Signal Processing and Communications Applications Conference (SIU), July 2018

[3] R. Santhoshkumar, M. Kalaiselvi Geetha" Recognition of Emotions from Human Activity Using STIP Feature",www.ijesi.org, International Journal of Engineering Science Invention, 2018, Pp.88-97

[4] B. Lavanya; G.S. Gayathri, "Exploration and Deduction of Sensor-Based Human Activity Recognition System of Smart-Phone Data" IEEE 2017 International Conference on Computational Intelligence and Computing Research (ICCIC), Nov 2018

[5] Song-Mi Lee; Sang Min Yoon; Heeryon Cho" Human activity recognition from accelerometer data using Convolutional Neural Network" 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), March 2017

[6] Fatemeh Noroozi; Dorota Kaminska; Ciprian Corneanu; Tomasz Sapinski; Sergio Escalera; Gholamreza Anbarjafari" Survey on Emotional Body Gesture Recognition, IEEE Transactions on Affective Computing, October 2019, Pp 1-1

[7] Edward Price; George Moore; Leo Galway; Mark Linden," Towards Mobile Cognitive Fatigue Assessment as Indicated by Physical, Social, Environmental, and Emotional Factors", IEEE Access vol7, August 2019, Pp 116465 - 116479

[8] Jiayu Shu; Mangtik Chiu; Pan Hui "Emotion Sensing for Mobile Computing", IEEE Communications Magazine, Vol 57, November 2019, Pp 84-90

[9] A.V. Zaboleeva-Zotova; A.S. Bobkov; V.L. Rozaliev; A.B. Petrovsky "Automated Identification of Human Emotions by Gestures and Poses", IEEE 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence, July 2014

[10] Mohanad Babiker; Othman O. Khalifa; Kyaw Kyaw Htike; Aisha Hassan; Muhamed Zaharadeen" Automated daily human activity recognition for video surveillance using neural network",2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), Nov. 2017

[11] Di Wu; Ling Shao "Silhouette Analysis-Based Action Recognition Via Exploiting Human Poses "IEEE Transactions on Circuits and Systems for Video Technology, Vol 23, No 2, Feb. 2013, Pp 236 – 243

[12] Milad Mohammad Taghi Zadeh; Maryam Imani; Babak Majidi" Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters", 2019 5th Conference on Knowledge-Based Engineering and Innovation (KBEI), June 2019

[13] Marko Radeta; Marco Maiocchi" Towards Automatic and Unobtrusive Recognition of Primary-Process Emotions in Body Postures, Interaction, December 2013

[14] M. Shamim Hossain; Ghulam Muhammad," An Emotion Recognition System for Mobile Applications" IEEE Access, Vol 5, Pp 2281 - 2287