International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Human Actions and Hand Gesture Recognition with Deep Learning

Bapireddygari Hema, J. Arokia Renjit

Abstract*: Over recent times, deep learning has been challenged extensively to automatically read and interpret characteristic features from large volumes of data. Human Action Recognition (HAR) has been experimented with variety of techniques like wearable devices, mobile devices etc., but they can cause unnecessary discomfort to people especially elderly and child. Since it is very vital to monitor the movements of elderly and children in unattended scenarios, thus, HAR is focused. A smart human action recognition method to automatically identify the human activities from skeletal joint motions and combines the competencies are focused. We can also intimate the near ones about the status of the people. Also, it is a low-cost method and has high accuracy. Thus, this provides a way to help the senior citizens and children from any kind of mishaps and health issues. Hand gesture recognition is also discussed along with human activities using deep learning.

Keywords: Deep Learning, Human Action Recognition, Skeletal images, spatial dependencies and temporal dependencies, Hand gesture recognition, Transfer learning, machine learning, Convolutional Neural Network (CNN), Human Computer Interaction (HCI), Hierarchical spatio-temporal model (HSTM)

I. INTRODUCTION

Human action recognition (HAR) method extracts the features of skeletal images of human activities by using depth sensor camera automatically. This approach can be applied to monitor the elderly people, children and their environment & suspicious objects and people can be detected which can be done by using deep learning approach.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Bapireddygari Hema*, P.G Scholar, Computer Science and Engineering, Jeppiaar Engineering College, Chennai, Tamilnadu, India.
DR.J.Arokia Renjith, Professor & Head, Computer Science and Engineering, Jeppiaar Engineering College, Chennai, Tamilnadu, India

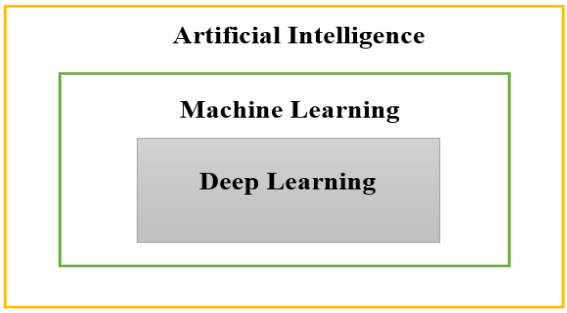

Fig. 1. Structure of Deep Learning

Deep Learning focuses on right features by their own with little programmer guidance. Deep learning models can deal with huge data, where machine learning cannot deal with high dimensionality of data. Deep learning is implemented through neural networks.

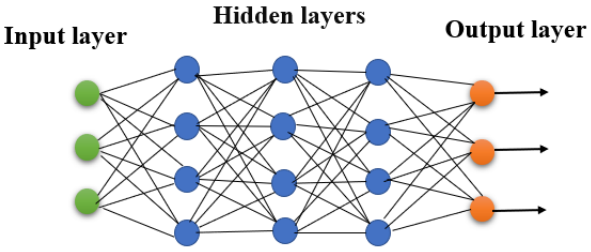

Fig. 2. Deep learning neural network

Human Action Recognition (HAR) is used in various applications. Detection of suspicious objects and people, monitoring elderly people while travelling or at home, monitoring the child left under someone or at home with full-time maid is becoming difficulty now-a-days. These situations will make the people worry to take care of their near and dear ones in problem. In the future or now, almost all the elderly people prefer to live in their own environment, and children are left under someone or at home with full-time maid as both the parents go for work. Affordable services with good quality are becoming hard to find. Therefore, automatic processing systems for monitoring the human actions in daily basis is playing an important role. Human action recognition method automatically identifies the human actions by skeletal joint motion images and hand gesture recognition is also included. This work is arranged as follows: Part II shows literature review of other works with the same concepts; Part III provides the challenges in Human action recognition; Part IV provides an overview of human action recognition and prediction;

Part V provides an overview of hand gesture recognition; Part VI provides datasets of different works from different authors; Part VII shows the results and discusses about the works of different authors related to the same concepts; Finally, Part VIII provides the conclusion.

II. LITERATURE REVIEW

In [1] author has focused on deep learning and image processing to show human action recognition system.
Author proposed a methodology for recognizing human daily activities by tracking the skeletal joint motions by using depth sensor camera. The proposed system of this work achieved 97% accuracy. In [2] author has spoken about Geometric dynamic configurations of body joints. This work uses a learning method which configures the automatic joints, by using sparse representation and dictionary learning. By using L0 norm constrained sparse coding and dictionary learning, the procedure automatically learns 4-D spatio-temporal features from body joint coordinate data. The performance is checked with three datasets such as state-of-the-art public HAR, with achievement of 94.12% accuracy which is better than other datasets performed. In human activity recognition, there are two key issues such as temporal and spatial dependencies. Only one of the methods is focused recently, thus complex activities cannot be recognized as there is no proper description. To solve such problem, [3] has proposed hierarchical spatio-temporal model (HSTM) which models temporal and spatial constraints simultaneously. To train all parameters, a novel learning algorithm is derived with bottom-to-up strategy. To obtain the superior classification ability, both the temporal and spatial similarities of actions are measured together. To recognize continuous human daily activities [4] has introduced an algorithm named as Fuzzy Segmentation and Recognition (FuzzySR) which expresses gradual transitions and their reasons. The main objective of the author is to recognize the activities in every event, to do so, a video is divided into number of events simultaneously. This algorithm solves the optimization problem and determines the most suitable activities label for every event. This work [5] uses transfer learning for classification of hand gestures. While leveraging the capacity of deep learning algorithms, transfer learning is applied on full data from several users. Three novel ConvNet architectures that are competitive with current sEMG-based classifiers are presented in this work. The new TL scheme that systematically and sufficiently improves the tested ConvNets performances are presented in this work. In Future research works, the TL algorithm on upper-extremity amputees will focus on adapting and testing. In [6] author introduces a sensor for recognizing hand gestures using impulse signals such as ultra-wideband that reflects from the hand, which is a wireless gesture recognition system. For gesture classification, machine learning alike convolutional neural network (CNN) is used. From the American Sign Language (ASL), Six hand gestures are experimented in this work and the accuracy result is greater than 90%.

III. CHALLENGES IN HUMAN ACTION RECOGNITION

The major challenges in human action recognition are as follows:

- Recognition of parallel activities: More than one action at the same time. Example- walking and talking with someone [8].
- Recognition of overlapped activities: Actions overlapping with each other [8]. Example- when he/she is cooking, and gets the phone call at the same time. Then he/she stops cooking for some time until they finish talking in the call.
- More than one occupant: Multiple occupants in one place. Example: Group of friends talking to each other in one place with different actions [8].
- Interpretation of actions: Similar activities interpreted in many ways. Example: Opening the fridge door to take vegetables or cleaning the refrigerator [8].
- Application domain: Based on this domain, the actions of details and interests will be varied. Example: Finding unusual behavior like jumping, falling down, Vomiting etc. [7].
- Inter and Intra class variations: The system performance is dependent on the wide range of variations of the activities. [7]. For example, Walking, jogging and running varies only by small degree.
- Paradigm usage learning: For recognizing different human activities, the learning-based approach is used. Based on the type of training data, the usage of learning paradigm may be either supervised or unsupervised [7].
- Recording and Background settings: Recognizing human actions with cluttered background is difficult. The system performance is decided based on the video quality. An efficient human action recognition must recognize the actions of human in different video quality and the cluttered background [7].

IV. HUMAN ACTION RECOGNITION AND PREDICTION

Human action recognition is the most popular research area in recognizing actions of human. Human action is defined as any specific behavior by human body [10]. Every human action is done for some purpose. Suppose, a patient in order to complete the physical exercise is interacting and responding using hands, legs, body etc., these actions are observed by eyes or by visual sensors. Through the human action it is easily understood that he/she is exercising. However, in real-world scenarios, like smart rehabilitation and visual surveillance, human labors are too expensive. So came an artificial intelligence, which builds a machine to understand the actions and intensions of humans accurately. For example, the patient undergoing rehabilitation exercise at his own place, then his/her robot assistant is able to recognize the actions, recognizes if the exercise done is correct or not and also prevents from further injuries.

Retrieval Number: B2815129219/2019©BEIESP
DOI: 10.35940/ijeat.B2815.129219
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Such type of intelligent machine is very beneficial and important as it saves time to visit therapist, medical cost is reduced and makes remote exercise into reality, which saves the labor cost as it is going to be very expensive in coming days [9].

There are two basic topics such as action recognition and action prediction.

Action Recognition: In computer vision community, recognizing actions is a fundamental task and human actions are recognized from a video containing complete action execution [9]

Action Prediction: It is a video understanding task before the fact and focusses on the future state. Human action is reasoned from temporarily incomplete video data.

The main difference between action recognition and prediction is to find "when to make a decision".

Human action recognition infers after the entire action execution is observed. Example: non-urgent scenarios such as video retrieval, entertainment, etc. Human action prediction infers before the entire action execution is observed. Example: A smart system in a vehicle which can analyze and predict accidents before it occurs [9].

The human action recognition framework is showed in fig 3.

In lower level, feature extraction and background subtraction are accomplished by pixel based and block-based approaches. Tracking and detection are performed by model based and feature based technique. In mid-level, actions are recognized after the tracking and detection. In high level reasoning engine interprets the human actions [7].

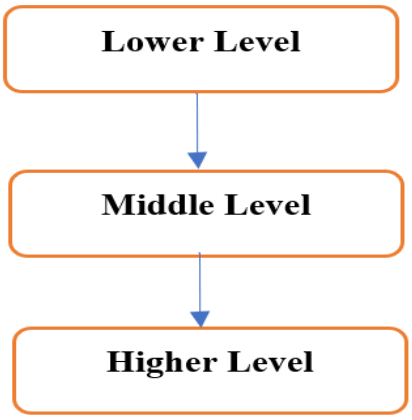

Fig. 3. Framework of Human Action Recognition

V. HAND GESTURE RECOGNITION

In recent years, Hand gesture recognition system became popular since it can interact directly with machine using human computer interaction (HCI) and also with its applications. To create a interaction between computer and human, recognition of hand gestures is built. For conveying meaningful information or controlling a robot, recognized gestures are used[11]. Human computer interaction (HCI) is also called as Man machine interaction (MMI). It refers to the connection between humans and computers. The two major characteristics are usability and functionality, that must be deemed while designing an HCI system. Set of Functions or services refers to the functionality, from system to users. In usability, the system can operate and perform by the level and scope. Gestures communicates between humans and machines and between human and human using the sign language. Gestures may be dynamic or static. Static such as certain pose or posture requires complexity to a smaller extent. Whereas dynamic postures are appropriate in real time environments but are more complex.

Gesture recognition system is mainly classified into three steps.

(i) Extraction method
(ii) Features estimation and extraction
(iii) Classification and recognition

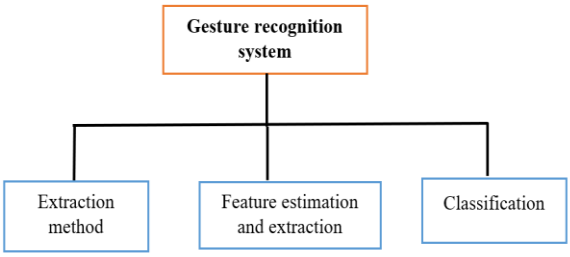

Fig. 4. Steps of Gesture recognition system

Hand movements of humans have controlling functions such as object grasping and object designing and also logically explainable functions such as sign language and pointing. Human skills belong to controlling the function. Teaching human skills to a system is a hard task, as it is difficult to describe manipulative hand movements[12].

The applications of recognizing hand gestures are: numbers recognition, Sign language translation, 3D modeling, television control, Virtual environments, smart surveillance, robot control, medical systems, Graphic editor control etc. [11].

VI. DATASETS

This section discusses about the datasets used by authors from literature review section.

In [1], author has proposed two datasets: the UTKinect Action-3D dataset, and the CAD-60 state-of-the-art public human-activity 3D dataset. Daily activities such as cooking, drinking water, phone answering is included in these datasets. In the UTKinect Action-3D dataset, same actions were inspected with different camera views, with different timings, and with different people in different way. In CAD-60 dataset, experiments are carried out with different environments, like companies, kitchen and bathrooms. In [2] author has proposed three datasets: Cornell Activity Dataset -60 (CAD-60) MSR Action3D, and MSR Daily Activity 3D. In [3], author proposed three types of datasets such as one-person action dataset (UCF), human-human interactional activity datasets (UT-Interaction, BIT-Interaction and CASIA), and human-object interaction dataset (Gupta video dataset). In [4] FuzzySR algorithm is proposed to accomplish human activity recognition and segmentation without interruption. To evaluate the performance datasets used are Hollywood-2, CAD-60, ACT42, and UTK-CAP. In [5], two datasets such as Myo dataset, and NinaPro DB5 are used. Author proposed TL algorithm to improve the performance. In [6] author has proposed wireless gesture recognition system with six gestures ("S", "E", "V", "W", "B", "C") in American sign language.

VII. RESULTS AND DISCUSSION

In [1] the results of datasets show the effectiveness of human action recognition proposed by the author. In UTKinect Action-3D dataset, accuracy is 97%, whereas in CAD-60 dataset, accuracy is 96.15%, precision 90.39%, recall 88.46%.

In the future, affordance hypothesis is taken into consideration to improve the system in action recognition. The author says that experiments on complex actions such as health-problems, like headaches and vomiting will be performed. In [2] among three datasets, CAD-60 dataset, gives accuracy of 94.12%, precision 90.18%, recall 92.86%. MSR Action 3D dataset, gives accuracy of 88.20%, MSR Daily Activity 3D gives 68.75% accuracy. The experimental results say that the method performs well with datasets and very simple. The performance is far better with CAD-60 dataset, than with MSR Daily Activity 3D dataset, because of noise and errors. In future, the performance can be improved by extending it to RGBD (Red Green Blue and Depth) images. In [3] the performance of HSTM has been checked on 3 tasks and with five datasets and found better results than the state-of-the-art methods. In [4] author proposed FuzzySR algorithm shows precision of 42.6%, 60.4%, 65.2%, and 78.9% on the Hollywood-2, CAD-60, ACT42, and UTK-CAP datasets, respectively. In [5], NinaPro DB5 dataset achieved 68.98% accuracy over ten participants. In future it will focus on testing and adapting the TL algorithm. In [6], results show that the proposed system achieved accuracy more than 90% with different gestures.

Table 1.Results of Existing methodologies

Author	Dataset	Accuracy	Precision	Recall
Cho Nilar Phyo , Student Member, IEEE, Thi Thi Zin, Member, IEEE and Pyke Tin	UTKINECT Action-3D CAD-60	97% - 96.15%	- - 90.39%	- - 88.46%
Jin Qi, Zhangjing Wang, Xiancheng Lin, and Chunming Li	CAD-60 MSR Action3D MSR Daily Activity3D	94.12% 88.20% 68.75%	90.18% - -	92.86% - -
Hao Zhang, Member, IEEE, Wenjun Zhou, Member, IEEE, and Lynne E. Parker, Fellow, IEEE	Hollywood-2 CAD-60 ACT42 UTK-CAP	- - - -	42.6% 60.4% 65.2% 78.9%	- - - -
Ulysse C ̂ot´e-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Cl´ement Gosselin, Kyrre Glette, Franc¸ois Laviolettey, and Benoit Gosseliny	NinaPro DB5	68.98%	-	-

VIII. CONCLUSION

This survey has showcased on human action recognition, hand gesture recognition and deep learning. Survey is done on different works with different technologies and methodologies with the concepts of recognizing hand gesture and human actions. The datasets used by particular authors are discussed. Challenges faced in recognizing human actions are showcased. The improved performance and accuracy are discussed with few review works. Now a days and near future recognizing the human actions in different scenarios plays major role to have a secure life. Recurrent neural networks and convolutional neural networks are the methods of deep learning which plays major role in identifying the features in depth and achieves state-of-the-art results, whereas machine learning cannot deal with the feature extraction, and high dimensionality of data. Thus, to recognize human actions and hand gestures deep learning is better than machine learning. Further research will improve the performance and accuracy more efficiently than the current existing methodologies by using AlexNet and VGG16 algorithms.

REFERENCES

1. Deep learning for recognizing human activities using motions of skeletal joints, Cho Nilar Phyo, Student Member, IEEE, Thi Thi Zin, Member, IEEE and Pyke Ti, IEEE 2018, Vol No: 0098-3063.
2. Learning Complex Spatio-Temporal Configurations of Body Joints for Online Activity Recognition Jin Qi, Zhangjing Wang, Xiancheng Lin, and Chunming Li, IEEE 2018, Vol No: 2168-2291.
3. A Hierarchical Spatio-Temporal Model for Human Activity Recognition Wanru Xu, Zhenjiang Miao, Member, IEEE, Xiao-Ping Zhang, Senior Member, IEEE, Yi Tian, vol:1520-9210, 2017
4. Fuzzy Temporal Segmentation and Probabilistic Recognition of Continuous Human Daily Activities Hao Zhang, Member, IEEE, Wenjun Zhou, Member, IEEE, and Lynne E. Parker, Fellow, IEEE, vol: 2168-2291, 2015.
5. Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning Ulysse C ̂ ot´e-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Cl´ement Gosselin, Kyrre Glette, Franc¸ois Laviolette†, and Benoit Gosselin, vol. 1534-4320, Mar. 2019.
6. A Hand Gesture Recognition Sensor Using Reflected Impulses, Seo Yul Kim, Hong Gul Han, Student Member IEEE, Jin Woo Kim, Sanghoon Lee, Senior Member IEEE and Tae Wook Kim, Senior Member IEEE, IEEE 2016, Vol No:1530-437X
7. A Survey on human activity recognition from videos. T. Subetha, Dr.S.Chitrakala, IEEE 2016, Conference paper.
8. Modelling and simulation of activities of daily living representing an older adult's behavior, Ahmad Lotfi, Abubaker Elbayoudi, 2015, conference paper.
9. Human Action Recognition and Prediction: A Survey, Yu Kong, Member, IEEE, and Yun Fu, Senior Member, IEEE, JOURNAL OF LATEX CLASS FILES, VOL. 13, NO. 9, SEPTEMBER 2018
10. A Survey on Human Action Recognition, Ayush Purohit *, Shardul Singh Chauhan*.
11. HAND GESTURE RECOGNITION: A LITERATURE REVIEW,Rafiqul Zaman Khan and Noor Adnan Ibraheem
12. Survey Paper on Hand Gesture Recognition, Manjunatha M B, Pradeep kumar B.P,Santhosh.S.Y

AUTHORS PROFILE

Bapireddygari Hema, is a PG Scholar at Jeppiaar Engineering College, Chennai, Tamilnadu. She received the B. Tech degree from Sri Venkateswara College of Engineering and Technology, Chittoor, Andhra Pradesh. Certified in cloud computing and Infosys campus connect program and also presented papers on palm vein technology and Cloud computing. Her area of interests are Deep Learning, Big Data and IOT.

Dr.J. Arokia Renjit, B.E, M.E, Ph.D., works as the professor and Head of the CSE department of Jeppiaar Engineering College, Chennai, TamilNadu, India. He has more than 17 years of teaching experience, and his areas of specialization include Data Mining, Artificial Intelligence, Image processing, Cryptography and Network Security. He has received a funding grant of ₹8.81Lakhs from Department of science and Technology under SERB scheme. He has published more than 20 research papers in reputed international journals and in the proceedings of National and International level Conferences. He has guided more than 20 student projects in undergraduate level and postgraduate level. He is currently guiding for 8 PhD scholars under Anna University.