Self-Intelligence with Human Activities Recognition based on Convolutional Neural Networks

R.Jayaraj, Karishma Agarwal, Utkarsh Singh, Aparna Singh

Abstract: ***In the presented paper, we propose a strategy related to activity recognition of human from profundity maps as well as sequences stance information using convolutional neural systems. Two information descriptors will be utilized for activity portrayal. The main information is a depth movement picture which will store back to back depth motion images of a human activity, whilst the subsequent data is the proposed moving joint description feature which conveys the movement of joints after time instants. To boost highlight extraction for precise activity arrangement, we will use three networked channels prepared with different inputs along with hypothesis verification. The activity results produced from those channels are intertwined for last activity characterization. Here, we suggest a few combination score based tasks to amplify the weightage of the correct activity. The experiments reveal the aftereffects of intertwining the yield of those channels along with the hypothesis are superior to utilizing a single channel or intertwining more than one channel in particular. The technique was assessed on two open databases which are Microsoft activity dataset and the second one is taken from University of Texas . The results demonstrate that our method beats the vast majority of existing cutting edge techniques, for example, histogram of arranged 4-D normal in datasets. Albeit DHA dataset has high number of activities (38 activities) contrasted with existing activity datasets, our paper outperforms a cutting edge strategy on the dataset by 6.9%.***

Keywords**: Convolutional neural networks (CNN), activity recognition, deep learning.**

I. INTRODUCTION

Convolution Neural Networks[1] are specialized linear operations that use convolution instead of usual matrix multiplication in at least one of their layers .Convolution layer mainly comprise of input, multiple hidden and output layers. CNN is based on the fact that the input has the image and constrains the design in more sensitive way, it has also improved the development of other machine learning approaches. It is a class of deep feed artificial network that has successfully been applied to analyzing visual imagery.Nowadays, human activity recognition is pertinent for various computer application that require information about human's actions, not limited to but for inspection for public safety, robotics etc.

In our proposed paper we have used CNN which can take in an image as input, assign weights to various objects in that image. CNN is a good replacement from existing deep learning algorithms and is more efficient because it reduces the number of features and hence complexity. It is a specific type of neural network that uses perceptron's, machine learning algorithm, for monitored learning, to analyze and process data. Traditionally ,videos-based action recognition methods are used which is mainly based on processing inputs of two dimensions red-green-blue color images by utilized classifier like K-nearest neighbor[2], Bayesian network etc. We already know there are many pocket friendly devices such as kinect available and impressive features are provided by motion maps and positions in order to represent the action of human. On the other note they also have some drawbacks on existing techniques. For building up multi-view depth maps dataset we have extracted a huge number of features so as to provide a distinguished representation of each human activity for classification. While extracting feature from multi-view ,it might be possible to have same actions from front-view, but have different from side view. For performing different functions we have expanded number of layers. There are over 70 layers for performing different complex function and to solve algorithms. Some of the common layers are indulged like dropout and dense layer. Dropout reduces the over-fitting in the data which improves the performance of the algorithm on the other hand dense layer feeds all the outputs from the predecessor to all its neurons, each neuron providing the single output to the next layer. It is basic layer in neural network which contains ten neurons. In the presented paper, we have used two descriptors in order to represent the actions of human, to show sequence of map and joint descriptor variable to show the body posture sequence.

It connects depth maps which capture the change in depth of motion. The MJD[3] can change the missing lateral-view with it's informative representation which has great effect on boosting the performance. One of the operation named as score fusion operation has been devised to do the appropriate action with the help of various convolutional neural network channels.

Revised Manuscript Received on April 25, 2020.

R. Jayraj, Assistant Professor (UG),Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

Karishma Agarwal, Student, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

Aparna Singh, Student, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

Utkarsh Singh, Student, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

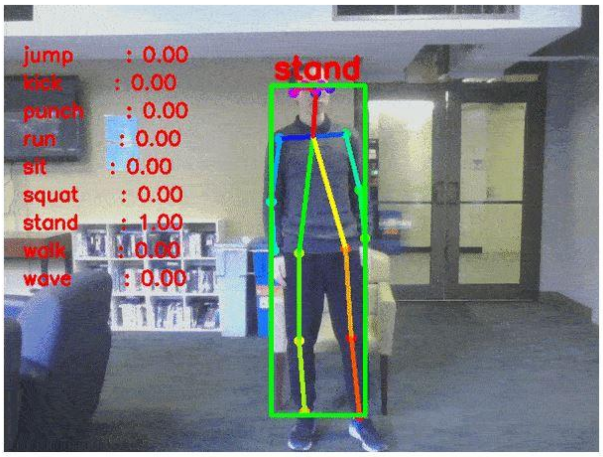

Fig 1.1 Above figure shows the matrix of human posture recognition of the person.

Our experiment suggested that while considering the highest score/marks of the three channel results it will lower the accuracy prediction on the tested data. The result expresses that the suggested method can identify human action more efficiently and with better and improved performance over the existing methods.

II. LITERATURE SURVEY

Nowadays the need for human interaction is getting popular in the field of robotic domain. Due to having a large no. of actions we need to have no. of features.

CNN is a class of deep learning networks mostly applied to analyze visual imagery. They are regularized versions of multilayer perceptron's. Multilayer perceptron's means fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. CNN take a different approach towards regularization which take advantage of the hierarchical pattern in data and assemble more complex pattern using small patterns which results in lower extreme complexity. Convolutional layer convolves the input and pass its result to the next layer which is similar to the response of a neuron in the visual cortex to a specific stimulus. CNN is a technique which is used for feature extraction and classification [4].

Depth based[5] approach provides an ordering from the center outwards such that the most central object gets the highest depth values and the least center objects the smallest depth. Skeletal based methods in parallel to depth- The given architecture shows the image processing classification of the input image. However, it is observe that the majority of CNN performance improvement came from redesigning of processing neurons and designing of new blocks. The input image is processed under various layers.Firstly, it is processed under convolution layer. In the based approaches [6]. Every joint is related to the local/subfield pattern description variable that provides highly discriminative features and translation invariant. For providing good 3-d activity recognition representation, a framework is proposed which is based on hypothesis generation and hypothesis testing along with matching. Zanfir[7] et al proposed non-parametric motion pose for low cache in human actions recognition, speed , movement of joints with respect to the current frame in the specific time window. Few proposed the action recognition to do in short films by sculpting them pyramidal and temperal design of human action images.[8].

III. SYSTEM ARCHITECTURE

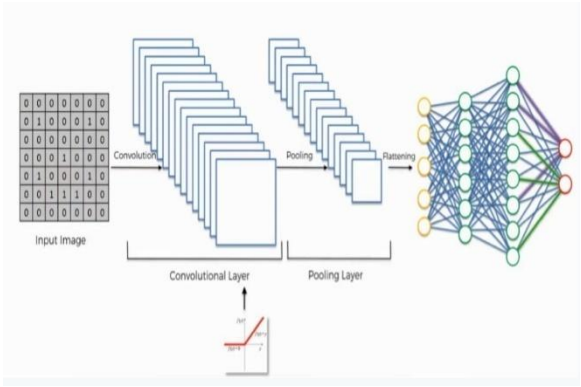

Fig. 1.2 System architecture

convolution layer it is processed according to the target output and then secondly, it is processed under polling layer in order to flatten the image according to the requirement in order to get the desired output [9].

IV. EXISTING SYSTEM

The existing system comprises of a technique for human activity acknowledgment from profundity maps and stance information three channels are superior to utilizing channel in particular. This proposed technique was assessed on two open databases which are Microsoft activity dataset and the University of Texas at Dallas-multimodal human activity dataset (DHA).The testing results demonstrate that the proposed approach beats a large portion of existing cutting edge strategies, for example, histogram of arranged 4D-neurons. In the above used RGB-D datasets the existing system take two datasets to get the performance model of the above given system [10]. The conclusion given by this system was that the position descriptor variable influence mostly the entire identification process to enhance first view for profundity figures representation, the combination experiments among the output predictors of the CNN channels will be used for maximizing the output step done. However, the results outperform most of the previous frameworks but it only yields good results in the test suite with still cameras from a pre-defined distance[11].

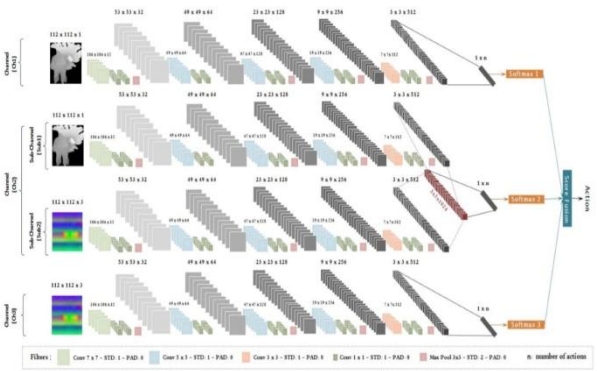

Fig 1.3 The above given figure shows graphical representation of the existing system which is having various channels of convolutional neural model for activity recognition

V. PROPOSED ALGORITHM

A.Data Preprocessing Step

Depth Image Processing-

The depth motion image describes the action which is overall reflected outwardly so as to produce an image which can be used to represent action with a particular outlook. The DMI will take into consideration the pictures captured by the device such as camera or mobile phone and will process each of the image containing body parts movement to full fledged action which further makes it easy for the model to extract the relevant information and perform manipulation.

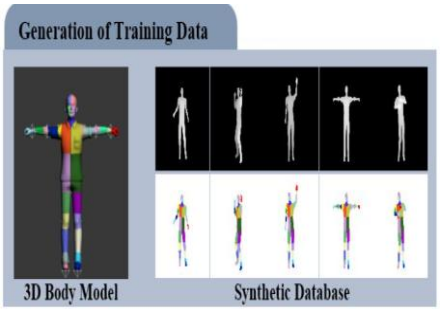

Fig 1.4 Generation of Training Data

- Moving Joint Descriptor:

The pictures captured by the smart devices, contain a huge number of joints. Out of those only some are useful in the context. We take cartesian coordinates which is sensitive to the joint movement and it is possible that it might represent same action for two or three different

B.Convolutional Neural Network Model

Model Description

The human recognition model is trained and learned based on convolution neural network of the deep learning, which has strong robustness to illumination difference, facial expression change and facial occlusion. The Convolutional Neural Network is a great achievement of Artificial Intelligence which can be used for image classification, object detection, semantic segmentation and human activity recognition. The Convolutional Neural Network (CNNs) is a bias of multi-layer perceptron motivated by biological vision and targeted at streamlining preprocessed data processes[12].Difference which is accounted between the two is that the cnn layer is made of convolution and the other is of pooling layer. The accuracy of the proposed system is largely depending on the many parameters. One of the main parameters is illumination condition[13]. The best conventional utilized histogram normalization procedure is histogram equalization where one tries to adjust the image histogram into a histogram that is persistent for all brightness values.

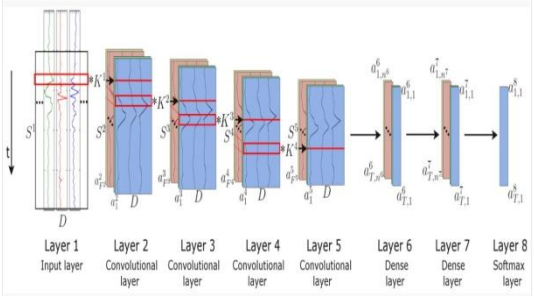

Fig 1.5 The above graph shows the depiction of CNN layer model for activity recognition.

The basic function is f(x) = y that describes the relation between x and y for all feasible inputs in the proper approach. The function f has to be decided from the hypothesis space [14].

Model Training

Model are processed under no. of multiple training channels. Each channel is represented by- a, b and c respectively .Channel 'a' is prepared with depth motion images, and the channel 'b' is proposed under image variables and MJD descriptor variables together, and channel 'c' is processed under MDR. The Channel 'b' is further sub-divided into two others sub-channels named as chn which is trained with descriptor as well as chn1 which is trained with MJD. After joining the last pooling layer it results in the formation of new one. The concatenation strategy was motivated by reference papers , which proposed different concatenation vide as which right forecast. In our test sets, numerous combination choices have been attempted, for example, component savvy averaging, greatest, expansion, and item, however the most extreme and item activities which we mean Max and Prod produce preferred outcomes over different tasks proposed an idea of joining the layers. Furthermore, the idea of joining the layers resulted in the better accuracy .After this initializationit resulted in an unstable and loss in the behavior. As already known, our learning supervised resulted in the loss and the decay of the weight which resulted in adecrease. We gave the input weights, then we already have trained the neural network model witha batch containing maximum 50 images. The looping or the iterations for every single channel's number is to get the minimum loss decay feature function.

Score Fusion

Each representation is having some information . In any case, for some tests, the most extreme worth doesn't speak to the right activity.

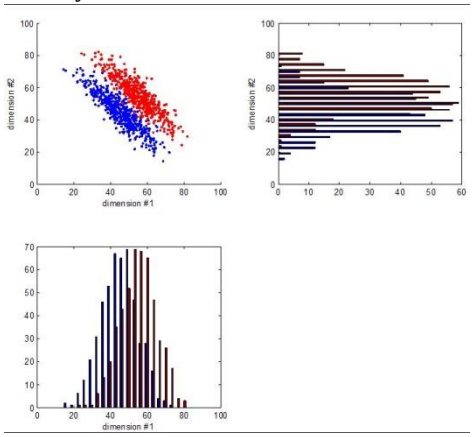

Fig 1.6 This diagram shows the score fusion graphical representation of human recognition.

A lower likelihood value than the most extreme may relate to the. As we will find in the exploratory outcomes area the grouping precision doesn't just relies upon the activity Max or Prod, however it additionally relies upon the directs engaged with the calculation.

VI. EXPERIMENT DESIGN

Here we extensively evaluate our proposed method on two public benchmark datasets namely MSRAction3D, and DHA. We employ convolutional neural network extreme learning machine (CNNELM) with occlusion and illumination performed on the dataset because of its better classification performance and efficient computation.

A.Benchmark Datasets

MSR 3D

The most widely used dataset used for the recognition of action is MSRAction3D. It has 20 actions: “high arm wave”, “horizontal arm wave”, “hammer”, “hand catch”, “forward punch”, “draw x”, “draw tick”, “draw circle”, “hand clap”, “two hand wave”, “bend”, “forward kick”, “side kick”, “jogging”, “tennis swing”, “tennis serve”, “golf swing”, “pick up &throw”. In the MSRAction3D dataset, actions such as “drawX” and “drawTick” are kind a of similar. One of the most challenging part of this dataset is mainly related to self-occlusion. Some pairs can be found in “leg-curl” and “leg-kick”, “run” and “walk”, etc. Second, our method directly uses depth motion maps with occlusions and illumination, which provide much better motion information.

DHA

We tend to used a chronic version of the DHA dataset wherever additional six action classes are concerned. [Lin et al., 2012] split depth sequences into reference system volume and developed 3bit binary patterns as depth options, that resulted in associate accuracy of eighty six.80% on the particular dataset. By incorporating multi-temporal info to the DMMs, our projected technique of exploitation occlusions and illumination achieves higher accuracy even on the extended DHA dataset. These enhancements show that operative datasets on multi-temporal DMMs will turn out a lot of informative options than operative on depth distinction motion history image (D-MHI).

B.Training and Testing time

The preparation time varies from a dataset to another, contingent upon the quantity of descriptors that are utilized for preparing. While MSRAction3D dataset has the least number of preparing information, the preparation time is additionally littler contrasted with the other two datasets that have more preparing information. We also notice that the preparation time also, number of emphases required for the model to unite are liable to number of preparing information. The instance of the DHA dataset is somewhat not quite the same as the other two datasets [15]. As the assessment convention of this dataset requests five preparing steps to ascertain the normal of the fivefold outcomes, the calculation preparing time for this dataset is the total of the five preparing spans. Despite the fact that the fivefold have a similar number of preparing information, the quantity of emphases required to get the base misfortune varies from one fold to another on the grounds that every datum has diverse mix of activities, and henceforth unique kinds of highlights to be educated. While the structure of the model utilized for preparing is the same for the three datasets just as the sort of preparing information. The equipment material utilized for testing and preparing is not the same as the one utilized for the preprocessing. For the most part, models require huge number of preparing models, for example, a huge instances pictures with long periods of preparing to arrive at a high expectation precision [16]. The key accomplishment of the learning process from a lot of information is to sufficiently separate highlights to perceive each activity.

VII. EXPERIMENTAL RESULTS

We have implemented the model in python using deep neural layers which turned out to improve the efficiency of the prediction. Output time was less than what we expected and the performance load was also reduced on the resources. Overall efficiency was eighty two percentage on both the datasets used. The models we have discussed constructs feature from spatial and temporal dimensions by implementing 3D convolutions. All the final feature representation is obtained by aggregating information from all channels. In the proposed paper we have considered the model for action recognition. Some deep architectures are also used such as deep belief networks which have promising performance on object recognition. In this paper, supervised algorithm is used to train developed 3D CNN and a large number of labeled samples are required. Earlier studies show that the number of labeled samples can be significantly lessen when such a model is pre-trained using unsupervised algorithms. A multi-temporal DMMs representation is suggested to seize more temporal motion information in depth sequence. Results shows that our method outrun the state-of-the-art methods in all datasets.

VIII. CONCLUSIONS AND FUTURE WORK

We developed 3D CNN models for action recognition. For the representation of a better action, two types of features has been proposed. The whole recognition process provides a great influence by providing parameters for front view of depth maps representation. As RGB-D dataset have small number of training samples. Our future work will focus on recognition of interconnection between people /objects.

REFERENCES

1. Vellingiri, J., S. Kaliraj, S. Satheeshkumar and T. Parthiban , "A Novel Approach for User recognition,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 47, no. 4, pp. 617–627, Apr. 2017.
2. S. Zhang, C. Gao, F. Chen, S. Luo, and N. Sang, “Group sparse-based mid-level representation for action recognition,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 47, no. 4, pp. 660– 672, Apr. 2017.
3. Y. Du,and L. Wang, “Skeleton based action recognition with convolutional neural network,” in Proc. IAPR Asian Conf. Pattern Recognit., 2015, pp. 579–583.
4. Y. Hou, Z. Li, P. Wang, and W. Li, “Skeleton optical spectra-based action recognition using convolutional neural networks,” IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 3, pp. 807–811, Mar. 2018.
5. J. Wang, Z. Liu, Y. Wu, and J. Yuan, “Mining actionlet ensemble for action recognition with depth cameras,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 1290–1297.
6. J. Zhang, W. Li, P. O. Ogunbona, P. Wang, and C. Tang, “RGB-Dbased action recognition datasets: A survey,” Pattern Recognit., vol. 60, pp. 86–105, 2016.
7. R.M.Rani, Dr.M. Pushpalatha," Generation of Frequent sensor epochs using efficient Parallel Distributed mining algorithm in large IOT", Computer Communications, Volume 148, 15 December 2019, Pages 107-114
8. R.Mythili, Revathi Venkataraman, T.Sai Raj,"An attribute-based lightweight cloud data access control using hypergraph structure", The Journal of Supercomputing(JoS),Published online: 02 Jan 2020 DOI: 10.1007/s11227-019-03119-7
9. S.Sivamohan, Liza.M.K, R.Veeramani, Krishnaveni.S,

Jothi.B, "Data Mining Techniques for DDOS Attack in Cloud Computing", IJCTA Interntional Scoience Press, Pg: 149-156
10. S Pandiaraj, Aishwarya, Surbhi, Alisha Minj, Priyanshu Singh, "Enabling Cloud Database Security Using Third Party Auditor", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-4, April, 2019
11. R.Veeramani,Dr.R.Madhan Mohan, "Iot Based Speech Recognition Controlled Car using Arduino", International Journal of Engineering and Advanced Technology,Volume-9 Issue-1, October 2019
12. T.H. Feiroz khan, N.Noor Alleema, Narendra Yadav, Sameer Mishra, Anshuman Shahi "Text Document Clustering using K-Means and Dbscan by using Machine Learning",International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019
13. S.Babeetha, B. Muruganantham, S. Ganesh Kumar, A. Murugan, "An enhanced kernel weighted collaborative recommended system to alleviate sparsity", International Journal of Electrical and Computer Engineering (IJECE), Volume 10, February 2020, Page No. 447-454
14. Kavitha.R ,K.Malathi,"Recognition and Classification of Diabetic Retinopathy utilizing Digital Fundus Image with Hybrid Algorithms", October 2019,International Journal of Emgineering & Advanced Technology(IJEAT), Volume 9, Isssue 1, 109-122
15. T.Chandraleka,Jayaraj R, " Hand Gesture Robot Car using ADXL 335 ", International Journal of Engineering and Advanced Technology (IJEAT)', Volume-8 Issue-4, Nov 2019
16. B.Sathya Bama,,Y.Bevish Jinila, "Attacks in Wireless sensor networks- A Research" ,International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8, Issue-9S2, July 2019
17. Vellingiri, J., S. Kaliraj, S. Satheeshkumar and T. Parthiban , "A Novel Approach for User Navigation Pattern Discovery and Analysis for Web Usage Mining", Journal of Computer Science 2015, vol 11 (2): Page no 372.382
18. C. Chen, R. Jafari, and N. Kehtarnavaz, “Action recognition from depth sequences using depth motion maps-based local binary patterns,” in Proc. IEEE Win. Conf. Appl. Comput. Vis., 2015, pp. 1092–1099.
19. J. Yu and J. Sun, “Multiactivity 3-D human pose tracking in incorporated motion model with transition bridges,” IEEE Trans. Syst., Man, Cybern., Syst., to be published.
20. W. Chi, J. Wang, and M. Q.-H. Meng, “A gait recognition method for human following in service robots,” IEEE Trans. Syst., Man, Cybern., Syst., to be published.
21. G. Liang, X. Lan, J. Wang, J. Wang, and N. Zheng, “A limb-based graphical model for human pose estimation,” IEEE Trans. Syst., Man, Cybern., Syst., vol. 48, no. 7, pp. 1080–1092, Jul. 2018. aterpillar: Confidential Green
22. Y. Guo, D. Tao, W. Liu, and J. Cheng, “Multiview Cauchy estimator feature embedding for depth and inertial sensor-based human action

AUTHORS PROFILE

R. Jayraj, Assistant Professor (UG),Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

Karishma Agarwal, final year undergraduate student, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

Aparna Singh, final year undergraduate student, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.

Utkarsh Singh, final year undergraduate student, Department of Information Technology, SRM Institute of Science and Technology, Ramapuram Campus, Chennai.