ISSN 2347 - 3983
Volume 7, No. 11 November 2019
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter227112019.pdf
https://doi.org/10.30534/ijeter/2019/227112019

Human activity recognition using Deep Learning

[1] M.V.D Prasad , Syed Inthiyaz [2] M. Teja kiran kumar[3] , K.H.S.Sharma[4],M. Gopi Manohar[5], Rupa Kumari[6], Sk Hasane Ahammad[7]
[1,2]Associate Professor, [3,7]Research Scholar, [3,4,5,6] B. Tech Students
[1,2,3,4,5,6,7]Department of ECE, Koneru Lakshmaiah Education Foundation, Guntur, India-522502

ABSTRACT

Presently now a day's productive human activity acknowledgment stays a difficult zone of research in the field of PC vision. Smarter reconnaissance is the need of this time through which ordinary and unusual exercises can be naturally distinguished utilizing fake information along with PC innovation. In our paperwork, we represented a structural movement acknowledgment in reconnaissance recordings caught upon mechanical frameworks. The constant observation running image stream is first isolated in significant snaps, in which images are chosen utilizing the analysed convolutional neural network-based human saliency highlights. Next, transient highlights of action in the succession of casings are separated by using the convolutional layers of a Flow Net to the CNN model. At long last, a multi-layer long transient memory (LSTM) is displayed for adapting large haul arrangements in the worldly optical stream highlights for action acknowledgment. The optical stream is determined, which gives the estimation of development, of A pixel in a succession of edges. Experiments are directed utilizing diverse activity and movement acknowledgment datasets and the outcomes uncover the viability of the proposed technique for action acknowledgment

Key words: action recognition, convolutional neural network (CNN), optical flow.

1. INTRODUCTION

human activities and occasions investigation in visual reconnaissance are fundamentally advantageous and have as of late advanced in the mechanical area, where several specialists on occupations need to be programmed observing. Human action acknowledgment is an interest of various mechanical applications, for example, video outline, savvy observation and checking frameworks, augmented reality, apply autonomy, restorative symptomatic, older social insurance, and substance-based video recovery. Human movement acknowledgment is the largest upcoming paperwork action in the PC vision field and has a great deal of viable application. Identifying action continuously naturally in open-air and indoor conditions request strength of highlights and classifiers. As the separation and point of the individual concerning camera may not be fixed, include extricated ought to be revolution invariant. The impediment is additionally a significant issue in packed situations like air terminal as the area of intrigue separated might be covering making extraction of shape highlights troublesome.

Programmed action acknowledgment is required for recognition of the anomalous movement in the air terminal, railroad station, transport stop and furthermore in the shopping centre. Action acknowledgment can be accomplished by separating the exercises dependent on its appearance, shape, intrigue point, optical stream or movement highlights. The (SVM) stands for support vector machine, the differentiator is one of the best-utilized classifiers for action classification.

2. LITERATURE SURVEY

In the current techniques, feature vectors, intrigue focuses or key points are separated. This progression has been skipped in Florence Simon paperwork. In Human action recognition using Optical flow [19] strategy, anticipating that it should lessen the intricacy in separating highlights, as these removed highlights could be scale variation or invariant. Yet rather whole edge is considered, for example every one of the pixels in the frame. Suspicion made is, starting and last casings in the video won't contain significant data as a result of elements like camera screen, henceforth those edges are not considered in figuring the optical stream. The distinction in direction and greatness are registered between the chose edges. Later these estimations of direction and greatness are set into canisters. The focused result, when contrasted with different strategies, is normal.

Mahesh Kumar H. Kolekar, in their paperwork [15], shape and optical flow features used together and used for the major work of human activity recognition. Highlights extracted are seen as proficient as a result given by the ANOVA test. Hidden Markov Model is produced for every action. The framework is prepared and tried in the different inside and open area conditions. The technique adjusted is made the outline and edge invariant. Precision accomplished utilizing least square support vector system is 80% for all exercises. Hidden Markov Model brought about better exactness when contrasted with the least square support vector machine classifier with the precision of 100% for strolling, 100% for

hand waving, 90% for bending, 84.61% for running and 90% for side gallop exercises. 100% exactness is accomplished in perceiving movement at the various point as for the camera. S Santhosh Kumar and mala john, in their paperwork [14], proposed a nearby descriptor worked by optical stream vectors along the edges of the activity performer. By utilizing the proposed highlight descriptor along with a multi-class Support Vector Machine classifier, acknowledgment rates as high as 95.69% and 94.62% have been accomplished for Weizmann activity set and KTH activity dataset respectively. The acknowledgment rate accomplished is 92.7% for UT connection Set_1, 90.21% for UT collaboration Set_2. The outcomes show that the technique is basic and proficient.

Amel Ben Mahjoub in their paper [12-18] presents their strategy to perceive human exercises. They utilize the spatiotemporal interest point for recognition of the significant difference in the picture. At that point, they extricate appearance and movement highlights of these intrigue focus utilizing the Histogram of Oriented Gradient (HOG) and Histogram Of Optical Flow (HOF) descriptors. At last, they coordinate the Support Vector Machine (SVM) by Bag of Word (BOW) of the space-time intrigue direct descriptor toward giving the name of every video grouping. They play out there a way to deal with the UTD-MHAD complex dataset and it gives a decent activity acknowledgment rate. The proposed calculation performs superior to anything different strategies dependent on similar succession information of the open UTD-MHAD database.

Two stream Model for our concept:

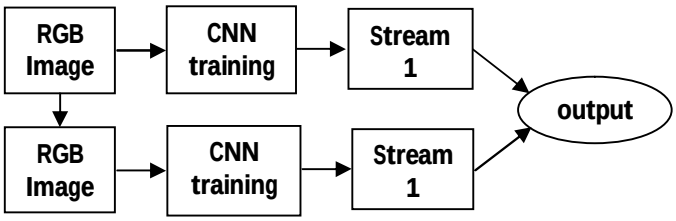

Figure 1: Flow diagram for training and feature extraction-classification on data.

From the figure 1 given above you can identify that our model is two stream CNN where we are taking the two process 1.RGB images and the other is optical flow pattern images which are given with certain colours with movement parts only. Where there are several two stream based models which includes different methods in two streams where they use different combinations in both blocks. Where in our model the patterned images obtained through the optical flow are processed through CNN and subtracted from the RGB sets and then the classification process takes places Experiment setup:-

In this section, From Fig.1 the experiment setup of human activity recognition is introduced. Especially, experiment scenarios, data collection, and detection of double hand waving , drinking water, walking and other actions are introduced.

Experiment Overview:-

As shown in Fig. 2, the experiment is about a human activity recognition of different actions like double hand waving, clapping, walking along with 20 other actions. As in this project we expect the system to recognize the activity of the system based on the data sets which we have trained through different algorithms.

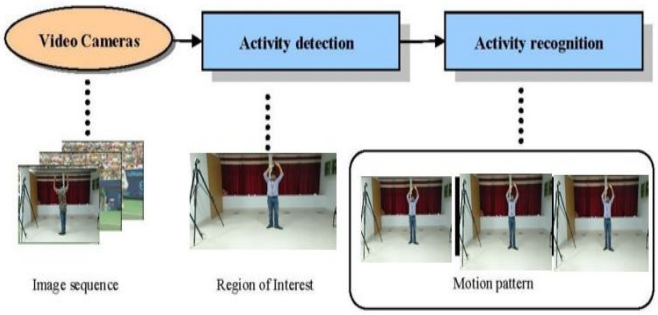

Figure 2: Representing The model of How the action are predicted using a camera.

As shown in Figure 2, the experiment is about a human activity recognition of different actions like double hand waving, clapping, walking along with 20 other actions. As in this project we expect the system to recognize the activity of the system based on the data sets which we have trained through different algorithms.

The workflow of the experiment is shown in the Fig. 3. Firstly we employ a camera to capture the motion of the human in the form of video. Next, we convert the video into RGB images and by using optical flow the motion of the human is detected and the colors are assigned to particular motions are used to train the deep learning-based prediction model. Finally, for every data or the video which we are going to give to the system we expect to recognize the action of the human-based on the data sets which we used to train the system.

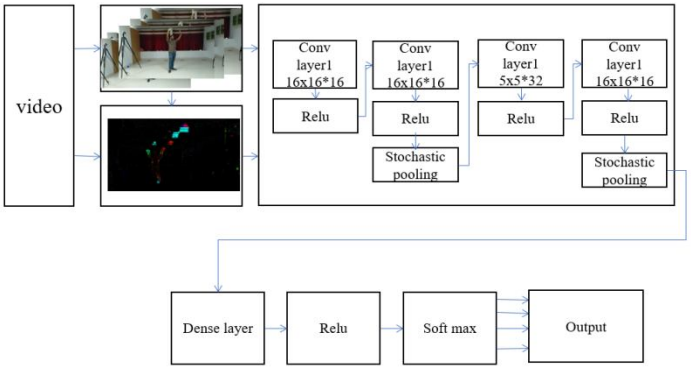

Figure 3: This Overall scheme represents the functioning of the model.

DATA COLLECTION

Figure 4: Different actions we took for the data base

From Fig.4 There are 4 human participants in this experiment. In this experiment we took 23 different types of actions and for every action we extracted 700 [RGB images and OPTICAL FLOW] and other images for testing. In total, we took 23 different actions and from those videos, we extracted 47000 testing and training images.

Data processing through Optical-flow

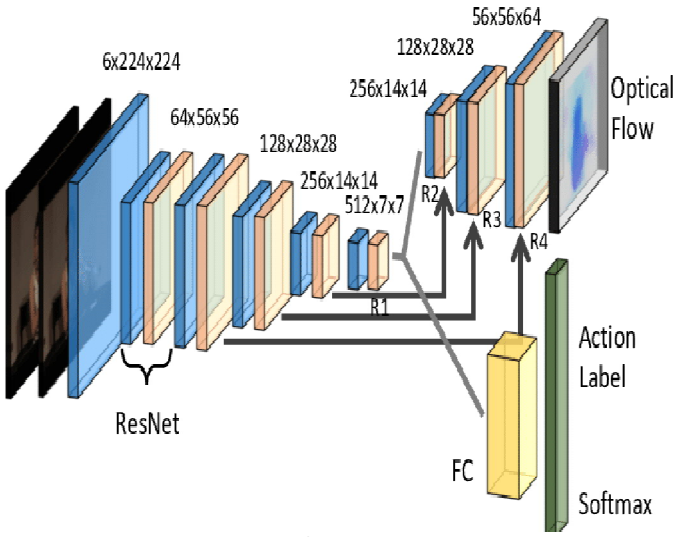

Figure 5: The Optical flow process is mentioned here.

Optical flow is the process where it generates patterns of visual objects and edges movements are tracked. From fig [5,6,7,8] we can see the pictures obtained through this process., optical flow is also defined as the patterns which are obtained through the velocities of the brightness of pattern motion in an image or activity has taken int. An American psychologist James placed the concept of optical flow. J Gibson identified the visual stimulus provided to animals moving around the world in the 1940s. Gibson emphasized the importance of optical flow, the ability to discern environmental action possibilities. Gibson's followers and his evolutionary approach to psychology further showed the role of the visual flow stimulus in the world's observer's perception of motion; perception of the form, length, and movement of objects in the world; and locomotion control.

These are images obtained through optical flow method:

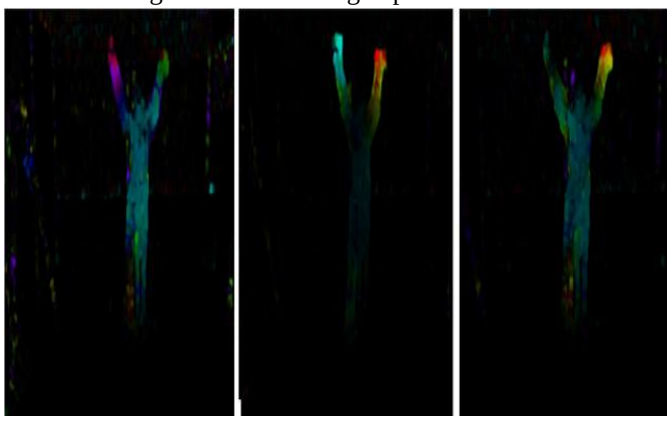

Figure 6: The optical flow data results for Double hand waving.

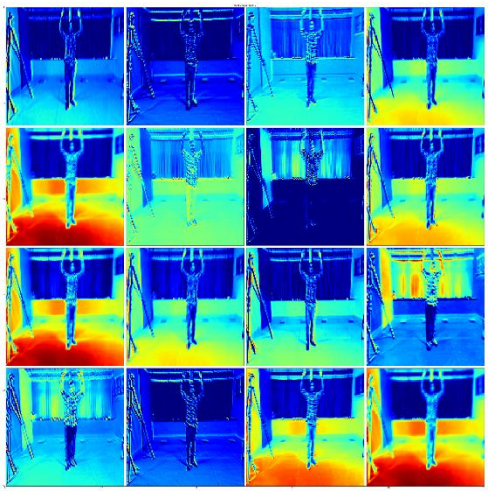

Figure 7: The optical flow data results for Double hand waving after feature extraction.

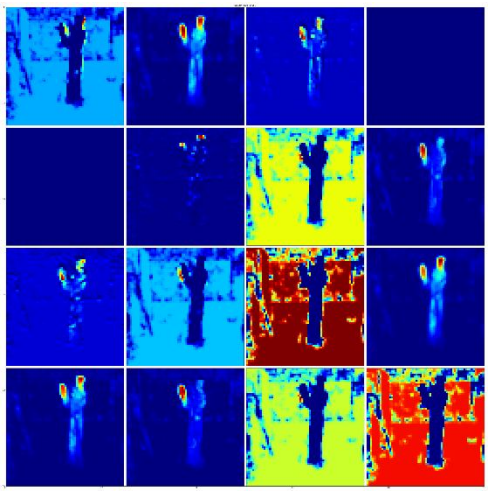

Figure 8: The optical flow data results for clapping after feature extraction.

The techniques of the optical stream try to ascertain the motion between two outlines of the picture that are now and again taken t and at each voxel location. Such techniques are called differential because they rely on the picture signal approximations of the nearby Taylor arrangement; that is, they use halfway subordinates about the spatial and worldly organizations.

Training

Here we are training our convolutional layers on 47000 raw data sets of human activities taken from a digital camera. It is based on more than 4 members and has a training set of 700 images of each activity along with the RGB training data set and optical flow data set. For pre-training, we utilize the first 23 convolutional layers pursued through normal pooling layer and completely associated layer. We train this network for around 1 day and accomplish a solitary field exactness of 89 percent on the understudy of human activity to detect by utilizing max pooling, then here we should convert or change the model which has to perform detection and in other models which is

included with combination of convolutional and associated layers to pre-trained networks can improve the execution.

3. METHODOLOGY

IMAGE ACQUISITION

image Acquisition is the main stage that circuits getting of corrupted leaf pictures to create a database. The RGB concealing pictures of people are taken from the advanced camera. The database of 30 pictures is gathered. During this development, we play out the maximum pooling strategy. From database five picture is taken and arranged further for testing

CONVOLUTION LAYER:

Convolution in image processing involves laying over another and then calculating a weighted sum of all pixel values. From the below Fig. [9,10] we can observe that the central pixel is replaced with a mutual mapped matrix so the 1x42 is replaced middle and then remaining all becomes zero in this manner we convolve images. Where in image processing convolution is nothing but the filtering process.

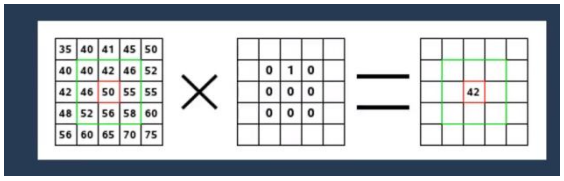

Figure 9: The Functioning of convolution layer.

The function used for Convolution layer.

$$x_j^l = f\left(\sum_{i \in M_j^{l-1}} x_i^{l-1} k_{ij}^l + b_j^l\right) - - - eq1$$

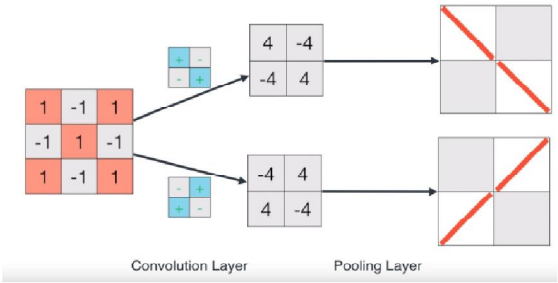

Figure 10: The working modality of convolution layers

From the above figure 10 you can observe the functioning of the convolution matrix and how it produces different models of images based on given pixel values.

MAX – POOLING

From Figure 11 we are seeing the functioning of max pooling and as you see the 4*4 matrix is taken and then the number of submatrices are divided from them which are based on desired qualities i.e. in below diagram the 2 grid is considered such that here the max element present in the subdivision of 2*2 matrix is taken u can observe from below block representation that from first 2*2 block " 8 " is selected because it is highest in given set similarly for all other subdivisions. The main purpose of max-pooling is to reduce the size of the image with unwanted data which takes more time to work so it reduces with the replacement of higher value in the set of pixel matrix required.

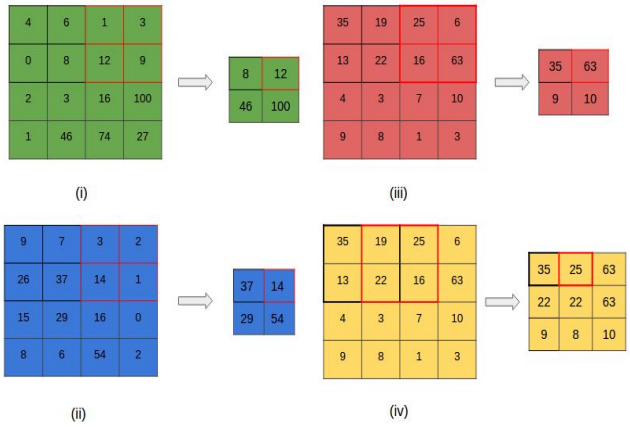

Figure 11: Max pooling block.

ACTIVATION FUNCTION

We need to check if the information received by the neuron for the given data is familiar or the same or should be ignored. It is known that the activation function is a non-linear transformation that we actually have to do over the input signal. Finally, the transforming output is sent as input data to the next level of neurons. From Fig.12 we can observe the functioning graph of ReLU.

$$y = activation\left(\sum(weights \times inputs) + bias\right)$$

There are many activation functions available, but we use ReLu Rectified Linear Unit, SoftMax function.

Rectified Linear Unit Layer (ReLu):

An amended direct unit has the yield as 0. We need to glance through whether the information is under 0 or not, and if the crude yield isn't suitable. Which implies, if the contribution of the direct unit is happened to be more prominent than 0 at that point, the yield clearly proportional to equivalent to include. ReLu s' apparatus is progressively similar to a genuine neuron in your body.

$$g(z) = max\{0, z\}$$

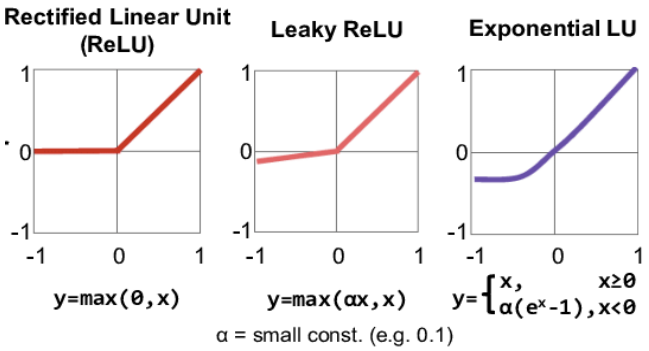

Figure 12: The functions for Relu

SoftMax Function

To calculate the probabilities of each objective of every class present over all viable objective classes by the use of SoftMax function. Later the probabilities of those classes which are classified and calculated will be helpful for verifying the objective class for the given inputs. The most important benefit of getting the output probabilities range is satisfied by the use of SoftMax function. The range is in between **0 to 1**, and the sum of all the probabilities considered to be **equal to one**.

$$P(y = j|\theta^{(i)}) = \frac{e^{\theta^{(i)}}}{\sum_{j=0}^{k} e^{\theta_k^{(i)}}}$$

$$where\ \theta = w_0 x_0 + w_1 x_1 + \cdots + w_0 x_k = \sum_{i=0}^{k} w_i x_i = w^T x$$

Results

classifiers	Recognition rate(%)					
	Batch-1 training		Batch-2 training		Batch-2 training	
	Testing with		Testing with		Testing with	
	Same data	Different data	Same data	Different data	Same data	Different data
ANN	76.39	74.26	78.47	75.29	79.13	77.42
Deep ANN	84.32	82.16	86.92	83.49	88.6	84.58
Convolutional Features and Multi-Layer LSTM	86.72	81.09	83.5	82.53	88.55	83.6
Our proposed method	92.34	89.92	90.34	88.69	89.62	89.34

Figure 13: The accuracy chart for proposed modal and other models

The result when compared to the other methods in the above-given chart we are getting more percentage of accuracy with our proposed model where the optical flow generated images are subtracted with another stream data i.e. RGB set data so by this we got good result with 91 % of accuracy. Where other models are having less than 90% and the no of matched models for given data and test data are also more. From Fig.[13,14,15,16] Those are the accuracy chart and the confusion matrix ,Training Accuracy vs Value accuracy for different batches.

FOR BATCH-1

Here we took the single activity where there are clean surroundings so we got 100% accuracy here.

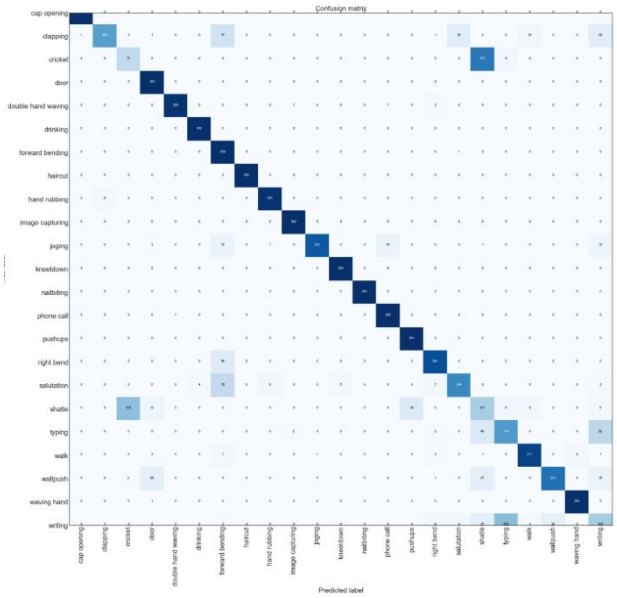

Figure 14: The confusion matrix for all 23 actions.

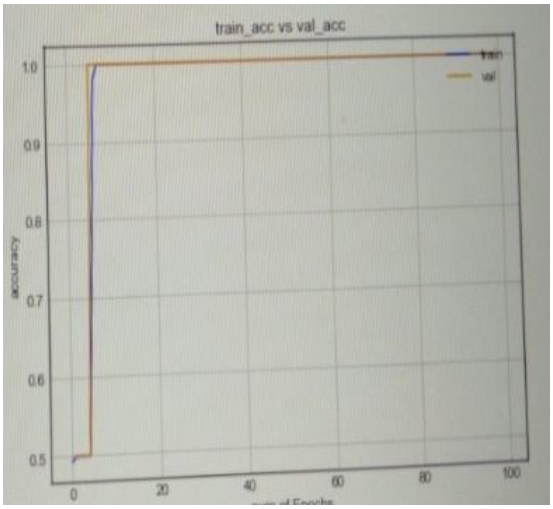

Figure 15: The taring accuracy vs value accuracy for 1st batch.

For the last batch2

Here the datasets of different more number of activities are taken so the accuracy is limited to 88%.

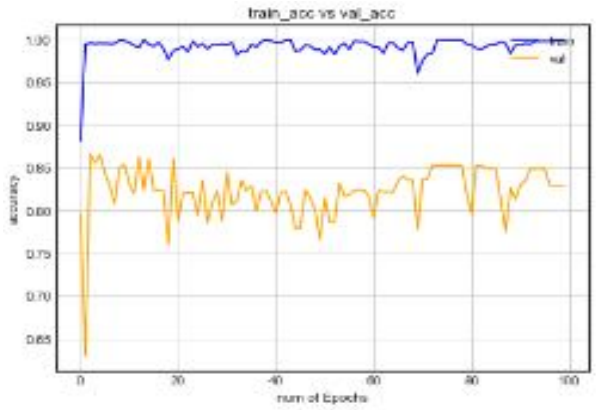

Figure 16: The taring accuracy vs value accuracy for 1st batch.

4. CONCLUSION

In our paper, we proposed a model which use full for video surveillance and also used in many different ways. where the concept is more useful for the robotics i.e. by training through algorithms the machines can easily recognize the human activities in the odd scenarios also i.e. In our proposed model the data sets are taken from very undesirable locations even though it's made in such way to detect those actions. So by this we will conclude that our proposed model will also work under very odd locations and places with very difficulties to detect the activities.

REFERENCES

1. Slim ABDELHEDI, Ali WALI, Adel M. ALIMI, **Human activity recognition based on mid-level representations in video surveillance applications.** Research Groups in Intelligent Machines, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia
2. Kyle Stephens and Adrian G. Bors," **Grouping Multi-Vector Streak Lines for Human Activity Identification**", Kyle Stephens and Adrian G. Bors Department of Computer Science, University of York, York YO10 5GH, UK
3. Li Liu, Ling Shao, Senior Member, IEEE, Xuelong Li, Fellow, IEEE, and Ke Lu **Learning Spatio-Temporal Representations for Action Recognition**: A Genetic Programming Approach
4. Zhigang Tu, Member, IEEE, Hongyan Li, Dejun Zhang, Member, IEEE, Justin Dauwels, Senior Member, IEEE Baoxin Li, Senior Member, IEEE and Junsong Yuan, Senior Member, IEEE **Action-Stage Emphasized Spatio-Temporal VLAD for Video Action Recognition**.
5. Qurat ul ain Ali Computer Engineering Department University of Engineering & Technology, **Role of Spatio-Temporal Feature Position in Recognition of Human Vehicle Interaction.**
6. Mrs. Savitha C Dr. D Ramesh Research Scholar Professor Sri Siddhartha Institute of Technology Tumakuru, India Tumakuru, India, **Motion Detection In Video Surviellance: A systematic Survey.**
7. Kyaw Kyaw Htike1 and Othman O. Khalifa1 1Electrical and Computer Engineering International Islamic University Malaysia, **Human Activity Recognition for Video Surveillance using Sequences of Postures.**
8. Neslihan K¨ose_ Mohammadreza Babaee_ Gerhard Rigoll Institute for Human-Machine Communication, TU Munich, Germany, **Multi-View Human Activity Recognition Using Motion Frequency.**
9. Debashis barman department of cse usha mary sharma, department of cse itit school of technology, assam don bosco university,Guwahati assam, India, **A study on human activity recognition from video.**
10. Congcong Liu, Jie Ying*, Feilong Han, Ming Ruan School of Optical Electrical and Computer Engineering University of Shanghai for Science and Technology Shanghai, China "**Abnormal Human Activity Recognition using Bayes Classifier and Convolutional Neural Network"**
11. Xiaoran Shi, Yaxin Li, Feng Zhou*, Lei Liu The Ministry Key Laboratory of Electronic Information Countermeasure and Simulation Xidian University Xi'an, Shaanxi, China "**Human Activity Recognition Based on Deep Learning Method"**
12. Amel Ben Mahjoub and Mohamed Atri Electronics and Micro-Electronics Laboratory Faculty of Sciences, Monastir, Tunisia Monastir University **Human action recognition using RGB data.**
13. Soumalya Sen, Moloy Dhar Susrut Banerjee Dept. of Computer Sc. & Engg. Dept. of Computer Sc.&Engg. Dept. of Computer Sc. & Engg Gurunanak Institute of Technology, **Implementation of Human Action Recognition using Image Parsing Techniques.**
14. Naresh Kumar Member, IEEE & Computer Society of India (CSI) Department of Mathematics Indian Institute of Technology Roorkee-247667, India, **Densely Sampled Noiseless Optical Flow for Motion based Visual Activity Analysis.**
15. Ahammad, SK Hasane, V. Rajesh, and MD Zia Ur Rahman. & quot; **Fast and Accurate Feature Extraction-Based Segmentation Framework for Spinal Cord Injury Severity Classification**. & quot; IEEE Access 7 (2019): 46092-46103. https://doi.org/10.1109/ACCESS.2019.2909583
16. Hasane Ahammad, S. K., & Rajesh, V. (2018). **Image processing based segmentation techniques for spinal cord in MRI**. Indian Journal of Public Health Research and Development, 9(6), 317-323 https://doi.org/10.5958/0976-5506.2018.00571.5
17. Ahammad, S. H., Rajesh, V., Saikumar, K., Jalakam, S., & Kumar, G. N. S. (2019). **Statistical analysis of spinal cord injury severity detection on high dimensional MRI data**. International Journal of Electrical and Computer Engineering, 9(5), 3457-3464.
18. Ahammad, S. H., Rajesh, V., Neetha, A., Sai Jeesmitha, B., & Srikanth, A. (2019). **Automatic segmentation of spinal cord diffusion MR images for disease location finding**. Indonesian Journal of Electrical Engineering and Computer Science, 15(3), 1313-1321.