HUMAN ACTION RECOGNITION SYSTEM

K. Gautam Naidu[1], P. Jagadeesh[2]

[1] K. Gautam Naidu, gautamnaidu713@gamil.com

UG Student, Saveetha School of Engineering, SIMATS, Chennai.

[2] P. Jagadeesh, jagadeeshp@saveetha.com

Assistant Professor, Saveetha School of Engineering, SIMATS, Chennai.

ABSTRACT- ***Human action identification is the major part of research area in computer science and its applications. The main purpose of the human action identification is to find out the ongoing occurrences and their related data from captured video. This technique is mainly used for monitoring sick persons and many numbers of systems that is directly connected with the computer interface system. Most of the real times application needs to recognize high level actions. The human actions are divided into three types based on their complexity. The first type is called as single performer actions. Walking and bending are the examples of single performer actions. The second type of human action is interaction. The interactions may be happened between human to human or human to object. Punching and lift the bags are the examples of interaction type. The third type of human action is group events. Dance with group members is one of the examples of group action. The camera was already installed in common places. The human action recognition is used to find the actions of the peoples to identify the important things. In this proposed system is used to identify the human action by using deep motion mapping concept and various computing algorithms. Finally the accuracy levels of the machine learning algorithms are compared.***

KEYWORDS- ***Human Action, Recognition, Depth motion map, RGBD camera, Classifier, Algorithms, machine learning***

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

I. INTRODUCTION

Human action identification is very important in this computerized world. The main objective of human action identification is automatically find out and succeeding events from unnamed video footage. This identification is used in various significant applications. Detection of human activities is used to identify normal and abnormal event in common places such as bus stand, railway station, hospitals, airports etc. Finding of changes in people's performance is significant task in action identification. Commonly two major action identification methods used are used to detect the events. The first method is to identify the action by using vision concept. The second method is to identify the activities by using depth map. In vision based method the various computer based methods are used to identify action of the people. The actions are captured by using camera or sensor devices. But in this method shows high variation between actual event and captured event. In the depth map method actions are identified by using features extracted from the images. In depth images are not having the texture feature. So it is very difficult to using various operations like gradients. The captured images are also too noisy. Calculating the accuracy level of the action recognition is tough in research point of view.

Depth images are captured from depth cameras. This camera is mostly used in animal monitoring concept. In this camera depth sensors are provide the picture for pick up positions of the animals and find out their activity and also it can sense the pictures with dark level. Depth cameras also used to detect the human skeleton by using depth map. This concept is applied in various applications like event identification and posture recognition.

This paper is arranged as; the author's fist describes the literature review and explains the important characteristics and problems that in related works. Section 3 shows the proposed work and outline diagram of the system. Section IV represents the results and discussion part. Section V concludes the proposed system.

II LITERATURE SURVEY

Vennila Megavannan et al., proposed a new method in this paper to identify people actions with the help of depth pictures. The characteristics of the pictures are collected by using bounding container using hierarchical part. Then the temporal data of the people action was collected from history of picture movements. Using the concept of translation identify the characteristics of motion. After that the pictures actions are classified by using SVM. Finally the efficiency level will be calculated for every feature. The end result shows the excellent performance when more than one features are integrated [1].

Chen Chen et al., presented a new method for identifying people actions. In this paper the authors used the concept depth motion pictures. The video part was projected by three planes. In each view two continuous map pictures are combined via video. Here collaborative classifier was applied with distance matrix. This method was developed and executed with real data. Finally the authors said that this method produce high performance compared with existing methods [2].

Heba Hamdy Ali a et al., says that human action identification by using depth Maps based is the major task of divide depth arrays with exact action. Most of the applications signify solutions in various areas like camera system, computerized application and video extracting model. This task is very difficult because the video setting may be different from one system to another system. In this paper the authors presented a proper study of latest depth map based picture presentation and character retrieve task. Here they discussed about the condition of data and classification task. The new identified method are applied in three important technique and higher accuracy result was collected [4].

S. Sandhya Rani et al., says that in current situation the various types of real time applications used the human activities recognition. The actual input data was collected from cameras from public places. The real time systems such as video extraction, people and robot communication and video coverage are mostly used the concept human activities recognition. Most of the techniques concentrate on few data only. This papers the authors presented a detailed survey about human action identification system. The review can be divided into two parts. In the fist part the authors presented various feature extraction concept by using classifiers. In the next part explained about every strategies of the human action recognition system [5].

Shugang Zhang et al identified the environment is the main factor to change the actions of the humans. In vision based research the inputs are collected from camera and human-system communication. In this paper the authors presented the review about human action identification. The important concept of this review was identifying action identification techniques and classification models. In the classification approach the authors categorized the methods based on template concept, models of discriminative and generative [6].

Linquin Cai et al., explains about the applications of human activities detection. It is used in different domains of computer field, pattern identification and human-system communication. In this paper the authors used modified CNN model to identify human activities. CNN model is one of the deep learning concepts. In this model initially extract the features of the images by using depth maps. In this modified CNN model 3D inputs and 3D task recognition is used to improve the computing speed and decrease the level of complexity of identification task. The final result shows that

this modified algorithm produce better result compare with other techniques [7].

Weiyao Xu et al., said that various sensors are provide the new chance for people activities by issuing image depth data. In this paper the authors proposed a new method for people activities detection using sampling method. Initially select sampling method was applied on the given data. Next static sampling method was used to find the sequences. Finally block based method is used to collect features from multilevel sampling. Then machine learning classifiers are applied the data set. The final result shows the performance of the proposed method with existing approaches [8].

Pichao Wang et al., said that in current situation deep learning approaches are mainly used in different domains. In this a new architecture was proposed to human activity identification. This architecture is called as Hierarchical deepness action Maps plus 3 direct Deep Convolutional Neural Networks. In this architecture unique depth information can be used. In the next stage extract the shape of the body and movement data. Finally the three planes are displayed separately [9].

Chen Chen et al presented efficient spatiotemporal used in human activity identification from video footage. In this proposed method the entire process can be divided into three parts. In the first part DMMs can be generated using captured pictures and motion. In the second part LBP was developed to identify the rotation of the image. In the last part apply various machine learning classifiers are applied to calculate the accuracy [10].

Enea Cippitelli et al., says that in surveillance, behavioral data analysis and the safety purposes the human action identification systems are used. In this research work the authors made a survey of video analysis using deep signal activity method.[11].

III PROPOSED METHOD

Human action identification is used in various places. The input data is collected from the cameras. The cameras are already installed in common places like railway station, bus stand, hospitals etc. The following fig1 shows that the flow diagram of proposed system

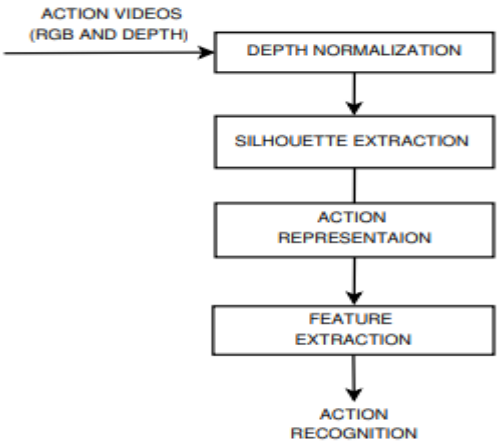


Figure 1 Outline of Proposed system

Here the task are divided into various steps. Initially the data is collected from the camera. After that the captured images are normalized by using depth value. In this normalization task two threshold data values are used in all activates. The following figure 2 shows the initial image from the camera and the equivalent normalized picture.

Figure 2 (a) Original Image (b) Normalized Image

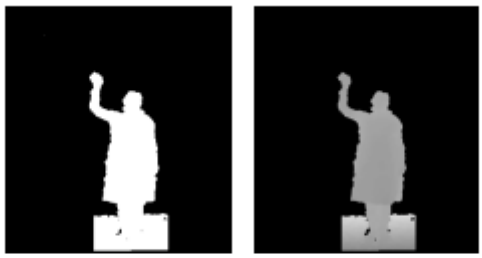
After the normalization the process, to delete the background by using depth picture. Here to find out silhouette image threhold values are used. Silhouette removal discovery and abolition of the surroundings using 2D RGB image was tedious task. Using depth picture the background portion of the image can be deleted easily. The threshold values are used to easily divide the background values. Finally silhouette of the depth picture can be created.

$$D(i,j,t) = \begin{cases} D'(i,j,t), & \text{if } D'(i,j,t) <= \zeta \\ 0, & \text{otherwise,} \end{cases}$$

Here i,j represents row and column value of the pixel and t is the time value of temporal frames, ζ represents threshold value D′ is the depth image finally D is the silhouette image from the original image.

$$B(i,j,t) = \begin{cases} 1, & \text{if } D(i,j,t) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

(i,j) specifies row and column value of the pixel location, t represents time value, mask B represent as the binary silhouette image D is the silhouette of the depth image. The following Figure 3 shows the binary (B) and depth (D) silhouettes images.



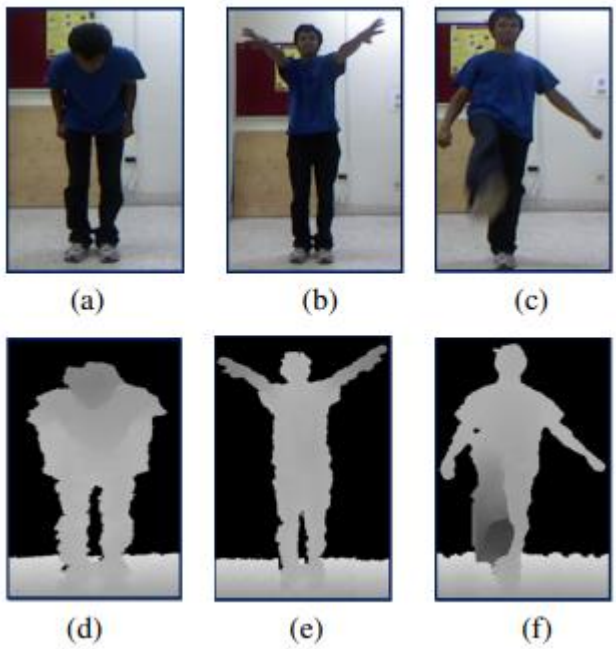
(a) (b)

Fig. 3. (a)Binary silhouette (b) Depth silhouette

In this proposed method the motion of the image is specified from history of an image. The motion history represents how long the image is moving. Average Depth picture to cover the movement in 3D (Z value).

IV RESULTS AND DISCUSSIONS

In this proposed system the images are captured from cameras. Then find out the depth normalization image. After that using threshold value generate Silhouette image from the normalized image. Finally the features of the images are extracted by using SVM approach. This section discuss about the database development, experimental arrangement and the output generate by SVM classifier concept. The following figure 4 (a-c) represents RGB images stored in the database and (d-f) represents the equivalent depth pictures.



(a) (b) (c)

(d) (e) (f)

Fig 5 Some of the actions considered in our database

(a) –(c) RGB images

Of bend, Swim, Kick (d) – (f) Corresponding Depth images

In this proposed system the database created by using Kinect. The RGB and original images are stored at the speed of 30 frames per second. The activities are collected 3 times and 350 to 400 frames of original pictures shows each activity per subject. The training set data contains the feature vectors of all the activities of all the pictures except one image. Here SVM classifier is used to classify the activities. The following table 1 shows the accuracy level. The accuracy level of Hu moments and hierarchical bounding box (HBB) can be displayed on the table.

Action	Recognition Accuracy (%)		
	Combined	Only HBB	Only Hu
Swimming	100	72.34	34.8
Bending	77.23	68.75	67.30
Waving	94.61	60.78	26.43
Stretching	81.23	43.37	43.37
Boxing	97.66	79.13	30.83
Bowling	97.94	93.80	90.0
Jumping	87.75	63.45	46.80
Kicking	83.33	59.37	25.30

Table 1 Accuracy Level

The data will be classified by using various machine level classifiers. The accuracy levels of the machine learning algorithms are shown in below.

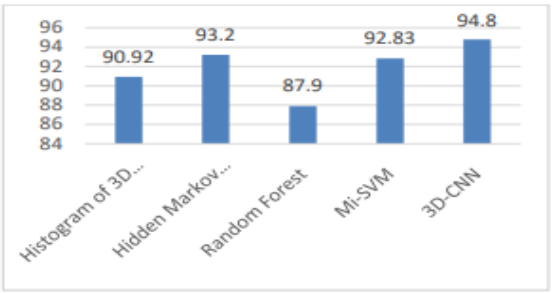


Figure 6 Accuracy of Various ML Algorithms.

The analysis show that 3D CNN machine learning concept provide better result compared with other machine learning concepts.

V CONCLUSION

In this proposed system human activities identification system using direct map concept. The original data is collected from camera. The cameras are already fixed in common places like bus stand, railway station etc. Initially the collected images from camera and convert to normalized images. Then Silhouette images are created and identify the action from the images. Finally the action images are given to input of machine learning algorithm. The accuracy levels of the machine learning algorithms are analyzed. This system is mainly used to identify the unusual conditions in the common locations

REFERENCES

[1] Vennila Megavannan Bhuvnesh Agarwal & R. Venkatesh Babu ,"Human Action Recognition using Depth Maps".

[2]Chen, C., Liu, K., & Kehtarnavaz, N. (2013), " Real-time human action recognition based on depth motion maps", Journal of Real-Time Image Processing, Vo.12, No. 1.

[3] Jiang Li; Xiaojuan Ban; Guang Yang; Yitong Li & Yu Wang(2019), " Real-time human action recognition using depth motion maps and convolutional neural networks", International Journal of High Performance Computing and Networking, Vol.13 No.3.

[4] Heba Hamdy Ali , Hossam M. Moftah & Aliaa A.A. Youssif(2018), " Depth-based human activity recognition: A comparative perspective study on feature extraction", Science Direct Future Computing and Informatics Journal, pp. 51 to 67.

[5] S. Sandhya Rani, G. Appa Rao Naidu & V. Usha Shree(2019)," A Fine Grainedresearch Over Human Action Recognition", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Vol. 9, No.1, pp 5376-5384.

[6] Shugang Zhang, Zhiqiang Wei, Jie Nie, Lei Huang,1Shuang Wang, & and Zhen Li(2017), "Review Article A Review on Human Activity Recognition Using Vision-Based Method ", Journal of Healthcare Engineering.

[7] Lingin Cai, Xiaolin, Fuli Chen & Min Xiang(2019)," Robust human action recognition based depth motion maps and improved Convolutional Neural Network".

[8] Weiyao Xu, Muging Wu & Xia Ting(2019), " Human Action Recognition Using Multilevel Depth Motion Maps",IEEE Access Computer Science.

[9] Pichao Wang, Wanqing Li, Zhimin Gao, Jing Zhang, Chang Tang, Philip Ogunbona," Deep Convolutional Neural Networks for Action Recognition Using Depth Map Sequences", Computer Vision and Pattern Recognition

[10] Chen Chen, Mengyuan Liu, Baochang Zhang, Jungong Han, Junjun Jiang & Hong Liu,) 3D Action Recognition Using Multi-temporal Depth Motion Maps and Fisher Vector ", Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence , pp 3331-1337.

[11] Enea Cippitelli, Ennio Gambi and Susanna Spinsante(2017),"Human Action Recognition with RGB-D Sensors", pp. 97-117.