

# Deep Learning Model for Human Activity Recognition in Aerial Images

Pooja M<sup>1</sup>, Jebah Jayakumar<sup>2</sup>

P.G. Student, Department of Computer Science and Engineering, BNM Institute of Technology, Bangalore, Karnataka, India<sup>1</sup>

Assistant Professor, Department of Computer Science and Engineering, BNM Institute of Technology Bangalore, Karnataka, India<sup>2</sup>

**ABSTRACT:** The Human Action Recognition (HAR) from images, video clips is motivated due to abundant availability of videos and images. There is a need of diverse applications for automatic observation of sick people, security of senior citizens and kids through interfaces between man and machine. Automatic labeling of activities of people in images captured through the cameras of closed circuit television will be the motive of such applications. In recent years there is huge demand of image analysis and in particular -tagging of suspicious actions in surveillance system. The main objective of this work is to develop a model that can identify and classify the predefined activities of an individual.

**KEYWORDS:** Deep Convolution Neural Network, Activity Recognition, RCNN.

## I. INTRODUCTION

The developments in artificial intelligence are more prominent in the reorganization of human action, especially in computer vision and pattern recognition. There are promising results in tagging the images and videos with the relevant actions. The process involves feature extraction from recorded images and videos based on relevance, developing model and identifying acts of human. Research is persuaded in developing programmed observations to detect suspicious activities in both public areas and private properties. The usage of such software can be in key areas for bus terminals, hospitals senior citizens in home, banks, office. It is really difficult to recognize human action from large gathering images and video clips, training the template and next finally recognition. In due process, it has to consider quality of images in terms of resolution, scene appearance, anthropometry which includes altitude & thickness, scale, camera position, motion, illumination, clutter in background and performance issues pose and execution rate. In the present approach the objective is to develop a machine learning prototype to recognize a sub set of human actions which are predefined. The training and testing is to be carried out on datasets which are considered as bench mark datasets. The technique used should also give emphasis on accuracy in classification, reduced computational cost and ability to provide reasonably good performance.

### 1.1 Applications

**1. Automated surveillance systems:** Daily life tracking systems usually aim to provide a simple guide for activity reporting, or to assist with exercise and healthy lifestyles. Such devices are fitted with embedded sensors such as accelerometer, gyroscope, GPS; monitoring steps taken by people, climbing stairs, burning calories, sleeping hours, moving distance, sleep quality, suspicious activities for real time reactions like fighting, stealing etc.

**2. Airports:** Through real-time cameras installed in drones and special security flights sensitive areas such as country borders, sea beaches can be identified for attacking enemies or engaging in terrorist or bomber aircraft entry.

**3. Radar Imaging systems:** Radar and sonar images are used to detect and recognize different target types or in aircraft or missile systems guidance. There is a multi camera system that uses a new calibration-free behavior recognition method for monitoring human activity at subway station.

**4.Intelligence analysis and gathering:** In sensitive areas like the border with the country, sea beaches can be identified and the tracing by means of real time cameras installed on drones and special safety flights by attacking enemies or engaging in terrorist or bomber activities.

## II. RELATED WORK

The problem of categorizing human activity has remained a challenging task in computer vision. [1]An approach to order the movement of human focus in a video clip. The entities which are moving are identified and their boundaries are removed. The inventors recommend utilizing a star skeleton conveyed through the entity's boundary as an aid for human movement examination. It has a star skeleton that gives two movement prompts: pose of the figure, and recurrent movement of skeleton fragments. These two factors plays important role in resolving the movements of exercise postures, for example jogging, running, and hand waving. It doesn't require an earlier human model, computationally affordable Some Issues 1.Creating relative joint images was difficult due to calculation of relative distance between referenced joints and the other joints.2.The joints in the frames where not within the predefined temporal dimension.[2] The point tracking approach is used when detected objects are represented as points. Tracking is performed by evaluating the objects state in terms of position and motion and by associating points across frames. The movement of an entity is to be keenly tracked to check is it in straight path, which can be trusted or dismissed as surrounding area. But in most applications the entities cannot be restricted to move in straight path. This limits the adoptability of the algorithm. Issues 1. It is expensive in terms of computational cost due its requirement. The computation is to be done for each and every point in stream which makes it less popular for applications in real time. 2.Requires external mechanism to detect in every frame. Problem areas in handling occlusion especially for objects represented by multiple points, misdetection, entries and exits of objects. [3] An approach to order the movement of human focus in a video clip. The entities which are moving are identified and their boundaries are removed. The inventors recommend utilizing a star skeleton conveyed through the entity's boundary as an aid for human movement examination.

## III. METHODOLOGY

The first stage of the proposed model is the collection of videos from the bench mark video data sets. i.e KTH dataset. These video clips collected are to be converted to frames. The next step is to determine the difference between the frames. This difference is used for the detection of activity. Once the object of interest has been detected, the background subtraction must be done. The final task is quality enhancement, which plays a vital role where RGB videos are low-quality videos. Human activity recognition involves actions such as running and jogging, etc., if the videos collected are low-quality videos, further activity recognition is a daunting task. Here R-CNN is used to train the data set videos to process the key features. In the proposed work we tend to initially recognize the activity performed by the person then eliminate the background pictures (by the technique of background subtraction), solely only specific person performing associate activity is known later the specific activity is labeled and recognition of performance takes place. In the validation step the output is validated using trained neural network with original and estimated values. . The important role in the entire process is the key point extraction which helps to categorize the action and is labelled accordingly.

**Step 1:** Load the video and convert in to frames

**Step 2:** Extract features

**Step3:** Train RCNN

**Step4:**Identify Activity

Pre-processing step in image processing also utilize Gaussian blur method, in order to improve the structures of the images at distinct scale, view scale place description and scale space application. Computationally, using a Gaussian smoothing to a frame is similar to convoluting the frame with a Gaussian-function. It is also called as 2 dimensional weier -strass transformation.

The RCNN algorithm propose a bunch of boxes in image.

**Step 1:** Takes video as an input.

**Step 2:** Generates initial segmentation so multiple regions from image is identified.

**Step 3:** Combines the similar group of regions to form larger region ( based on shape similarity, size similarity, color similarity)

**Step 4:** Finally these region proceed to find activity detection

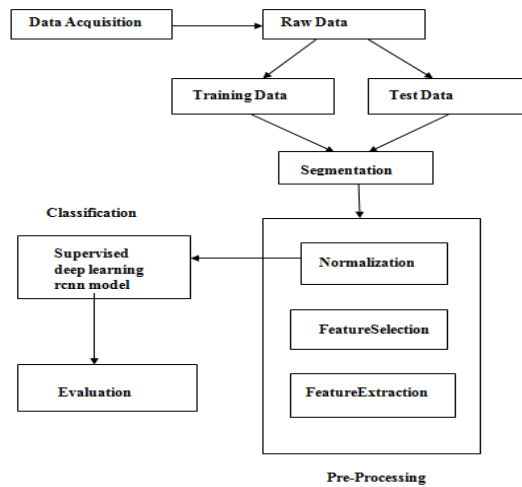


Figure 1-Proposed System for Human Activity Recognition

#### IV. EXPERIMENTAL RESULTS



Figure -2: Labelled activity recognized Person walking

The project is a python implementation. The Python program is run from the Idle IDE. There are different acts in the videos. First, video is collected from the directory, and then processed where the background containing the unwanted images is removed, and then processing steps are taken to identify activity only by people being moved for activity recognition.



The table 1 gives the detailed description about all the possible initial test cases which are required for the project

Test case	Test case description	Input	Output	Test case result
1	Uploading Video	Video dataset	Processing video frames	Pass
2	Processing of video	Video frames	Recognition activity from	Pass
3	Motion Detection	frames	Classifying	Pass
4	Recognition activity	frames	Activity recognized as walking, running etc	Pass

#### V. CONCLUSION

In this paper an accurate and efficient activity identification system has been developed which achieves comparable metrics with the existing state-of-the-art system. This paper uses recent techniques in the field of computer vision and deep learning. An important scope would be to train the system on a video sequence for usage in tracking applications. Addition of a temporally consistent network would enable smooth detection and more optimal than per-frame detection. Limitations are when there are limiting conditions, which means the project works only in certain criteria. This project cannot give an ideal, 100% accuracy, which is a limitation that comes naturally. The Model can be trained to make the classification multi-labelled classification. The deep CNN (Convolutional Neural Networks) classifiers are pre-trained. The efficiency of a RCNN classifier is more when compared to other neural network classifiers

#### REFERENCES

- [1] Sehar Un Nisa , Muhammad Imran , Shaheed Zulfikar, “A Critical Review of Object Detection Using Neural Networks”, International Conference on Communication computing and Digital systems, Bhutto Institute of science and technology Islambad, pp 154-159 ,2019.
- [2] Dangwei Li, Student Member, IEEE, Zhang Xiaotang Che,and Kaiqi Huang,Senior Member, IEEE “A Richly Annotated Pedestrian Dataset for Person Retrieval in Real Surveillance Scenarios”, University of Chinese Academy of Sciences, Beijing, pp 1-15 VOL. XX, NO. X, OCTOBER 2018
- [3] Russell Stewart, Mykhaylo Andriluka, and Andrew Y. Ng,“End-to-end people detection in crowded scenes”,Stanford University, USA Max Planck Institute for Informatics, pp 2325-2333,VOL III 2016.
- [4] Siji Joseph, Arun Pradeep “Object Tracking using HOG and SVM” University of Massachusetts Amherst Volume 48 ,pp 321-328, 6 June 2017.
- [5] Michael Rauter,“Reliable Human Detection in Top-View Depth Images” Austrian Institute of Technology Vienna, Austria,volume 2,pp 529-534,8 April 2017.