**ACM Pre-Internship**

# **Human Activity Recognition (HAR) for video surveillance**

Dr.Namitha K Girish S Anuvind M P R S Harish Kumar Harishankar Binu Nair

# **Research Project Report**

## **Introduction**

Human Activity Recognition (HAR) is a pattern recognition task that learns to identify human physical activities recorded by cameras and different sensor modalities. Human Activity Recognition is a major task in major fields such as computer vision and Artificial Intelligence. The importance of Human Activity Recognition (HAR) lies in its ability to monitor and track human activities in real-time, which can provide valuable insights and support for several applications.

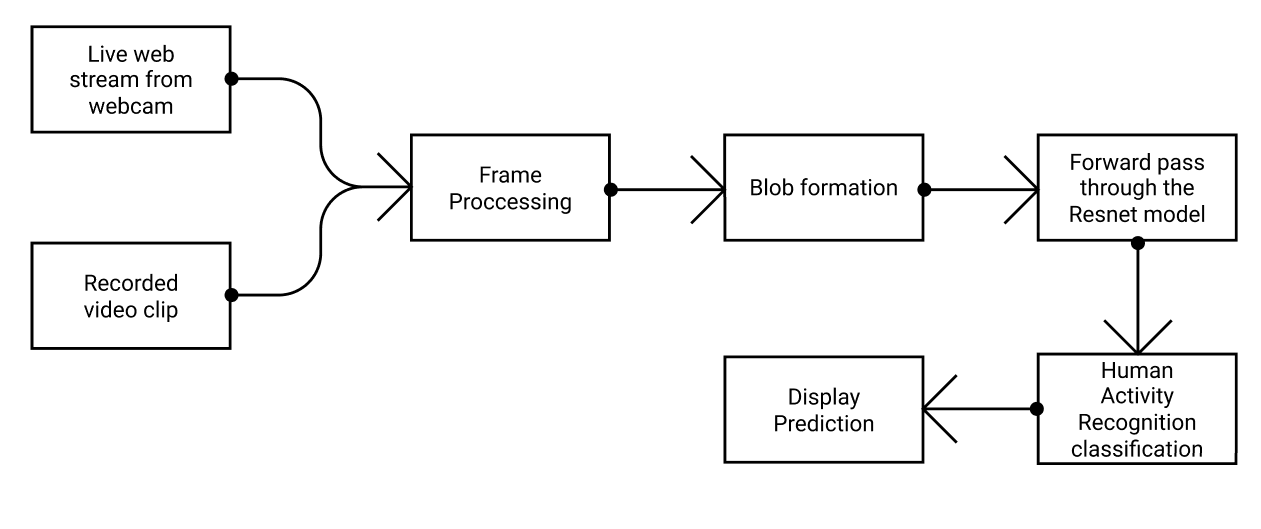
HAR systems have applications in various areas, content-based video summarization, human–computer interaction, education, healthcare, video surveillance, abnormal activity detection, sports, and entertainment. HAR methods are composed of three important components: (1) video frame segmentation for action detection, (2) action representation with respect to posture and motion of the human body, and (3) learning process that recognizes these actions.

The relevance of this problem is widespread, with applications ranging from video surveillance and security systems to human-computer interaction and sports analytics. The accurate recognition of human activities is crucial in these domains for real-time decision-making and behavioral analysis. Multiple open challenges exist within this problem space, including robust activity recognition in varying environmental conditions, real-time processing capabilities, and the efficient handling of extensive datasets. In this research, we intend to explore pre-existing research in the field and primarily focus on enhancing real-time recognition performance in dynamic scenarios.

## **Literature Survey**

The survey has provided valuable insights on methods to implement HAR and increase proficiency in the field to increase accuracy. The survey conducted on 8 different papers which contains various methods such as background subtraction, recognition on static and dynamic activity separately, pose estimation, data enhancement before classification, HOG descriptor, tube extraction, binary image creation for human edge detection, convLSTM for classification, linear SVM classifier, etc.

## **Implementation**

A resnet kinetic ONNX model was taken to classify human activity. In this implementation either video or live webcam can be classified. From the video stream frames are extracted and appended to a Double Ended Queue from collections package, the Queue contains space for 16 frames at a time. Then the sequence of images stored in the Double Ended Queue is processed and converted into an image blob using blobfromimages() method from cv2 . This image blob is then processed and its dimensions changed to fit as the input for the resnet model, a new batch dimension is added for batch normalization, the frames are resized and the dimensions are reordered to fit the input for the model. Model takes an input dimension of (1, 16, 112, 550, 400, 3)

## **Result**

The model’s validation provided valuable insights on the performance, it has successfully classified the test data to some extent. But, there exists hints of overfitting, thus we take a new approach in the phase 2 to create a new model from scratch and train on a lower number of action classes, the current model has 400 classes trained. In the next phase, it is planned to train on just 7 classes and use data preprocessing techniques to increase the accuracy of the model.