International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

Recurrent Neural Network for Human Action Recognition using Star Skeletonization

Nithin Paulose[1], M. Muthukumar[2], S. Swathi[3], M. Vignesh[4]

[1,2]B.E.Computer Science and Engineering, Dhaanish Ahmed Institute of Technology, Coimbatore, Tamil Nadu, India
[3,4]Assistant Professor, Dept. of Computer Science and Engineering, Dhaanish Ahmed Institute of Technology, Coimbatore, Tamil Nadu, India

---

Abstract - *This project presents a Recurrent Neural Network (RNN) methodology for Human Action Recognition using star skeleton as a representative descriptor of human posture. Star skeleton is a fast skeletonization technique by connecting from geometric center of target object to contour extremes. For the action recognition using the feature star skeleton, we clearly define the feature as a five-dimensional vector in star fashion because the head and four limbs are usually local extremes of human shape. In our project we assumed an action is composed of a series of star skeletons overtime. Therefore, the images which are time-sequential are expressing human action that is transformed into a feature vector sequence. Then the feature vector sequence must be transformed into symbol sequence so that RNN can model the action. We used RNN because the features extracted are time dependent.*

Key Words: *Recurrent Neural Network (RNN), star skeleton, contour extremes, Human Action Recognition, five-dimensional vector, time-sequential.*

1. INTRODUCTION

Human activity recognition is an important task for Ambient Intelligence systems. The state of a person is to be recognized, which provides us with valuable information that is been used as input for other systems. For example, in health care, fall detection can be used to alert the medical in case of an accident; in security, abnormal behavior can be detected and thus used to prevent a burglary or other criminal activities. Human motion analysis is currently receiving increasing attention from computer vision researchers. For example, Human body segmentation in an image, the movement of joints are tracked in an image sequence, and the analysis of athletic performance is done by recovering underlying 3D body structure , also used for medical diagnostics. Other applications include building man-machine user interfaces and video conferencing.

The goal of human activity recognition is to automatically analyses on -going activities from an unknown video. The objective of the system is to correctly classify the video into its activity category, for example where a video is segmented to contain only one execution of a human activity. The starting and ending times of all occurring activities from an input video is detected, from which the continuous recognition of human activities are performed. The constructions of several important applications are constructed from the videos which has the ability to recognize complex human activities. Automated surveillance systems in public places like airports and subway stations require detection of abnormal and suspicious activities, as opposed to normal activities. For example, an automatic recognize of suspicious activities like “a person leaves a bag” or “a person places his/her bag in a trash bin” in an airport surveillance system must be recognized. Using recognition of human activity the real-time monitoring of patients, children, and elderly persons can be done. By using activity recognition the construction of gesture-based human computer interfaces and vision-based intelligent environments becomes possible.

There are various types of human activities. Depending on their complexity, they can be conceptually categorized into four different levels: gestures, actions, interactions, and group activities.

1.1 HUMAN ACTIVITY RECOGNITION FROM VIDEO SEQUENCES

Human activity recognition role is that human-to-human interaction and interpersonal relations. HAR provides information about the identity, personality, and psychological state that is difficult to extract. In various classification techniques two main questions emerge: “Where it is in the video?” and “What is the action?” To recognize human activities one must determine the active states of a person, to recognize the efficiency. “Walking” is the daily human activity to recognize. The complex activities such as “peeling an apple” are more difficult to identify. The easier way to recognize is to simplify the complex activities into other simpler activities. For the better understanding of human activities the detection of objects may provide useful information about the ongoing event.

The human activity recognition assumes a figure-centric scene of uncluttered background, where the actor is free to perform an activity. The challenging task to classify a person’s activities with low error in a fully automated human activity recognition system are background clutter, partial occlusion, changes in scale, viewpoint, lighting and appearance, and frame resolution. Moreover, commenting

behavioral roles is time consuming and requires knowledge of the specific event. However, intra- and inter-class similarities make the problem more challenging, that is ,the same action done by two people may finding difficult to detect their individual action . The human activity is based on their habits, and this make s so difficult to determine their activity. In real time the challenging tasks is that, the construction of visual model for learning and analyzing human movements with inadequate benchmark datasets for evaluation.

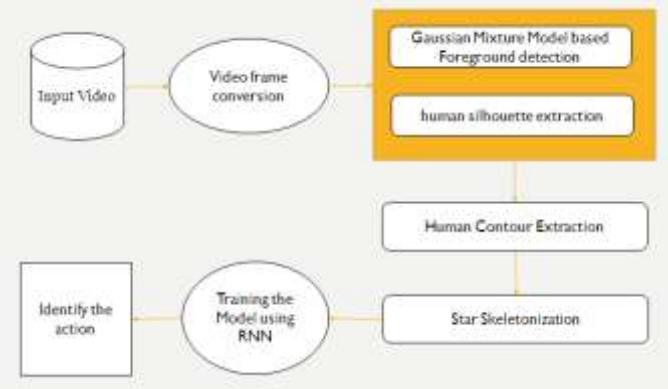

Fig 1: Architecture diagram

1.2 HUMAN ACTIVITY CATEGORIZATION

The human activity recognition methods are classified into two main categories: (i) uni-modal and (ii) multimodal activity recognition methods according to the nature of sensor data they employ. Then, these two categories is further break down into sub-categories depending on how they model human activities.

Uni-modal methods represent human activities from data of a single modality, such as images, and they are further categorized as: (i) space-time, (ii) stochastic, (iii) rule-based, and (iv) shape based methods. Space-time methods involve activity recognition methods, which represent human activities as a set of spatio-temporal features or trajectories. By applying statistical models to represent human actions stochastic methods are used. The modelings the motion of human body parts are shape-based methods efficiently represent activities with high level reasoning.

Multimodal methods combine features collected from different sources and are classified into three categories: (i) affective, (ii) behavioral, and (iii) social networking methods. Emotional communications and the affective state of a person represent human activity. Behavioral methods aim to recognize behavioral attributes, non-verbal multimodal cues such as gestures, facial expressions and auditory cues. The characteristics are model by social networking methods .It model the characteristics and the behavior of humans in several layers of human-to-human interactions in social events from gestures, body motion, and speech. The activity and behavior are the two terms which are used interchangeably. These two terms activity and behavior is used to describe a sequence of actions that correspond to specific body motion and to characterize both activities and events that are associated with facial expression, emotional states, and gestures with single person auditory cues.

1.3 HUMAN ACTIVITY RECOGNITION MODEL

The human activity recognition has largely focused on statistical methods using spatio-temporal features. The typical model consists of spatio-temporal interest-points which are detected in the video sequence and the local maxima become the center point of a spatio-temporal region. Features are then extracted from the spatio-temporal region (such as features based on optical flow or gradient values) and summarized or histogram med to form a feature descriptor. The feature descriptors are used to form a code book, typically followed by a bag of visual words model adapted from statistical natural language processing. While methods based on spatio-temporal features are the most common, other methods make use of other video features such as medium-term tracking, volumetric representations and graph-based features.

2. LITERATURE REVIEW

[1] A computational efficient action recognition framework using depth motion maps (DMMs)-based local binary patterns (LBPs) and kernel-based extreme learning machine (KELM). In depth video sequence depth frames are projected onto three orthogonal Cartesian planes to form the projected images corresponding to three projection views [front (f), side (s), and top (t) views].For calculating LBP histogram an LBP operator is applied to each block and the DMMs are divided into overlapped blocks. Feature-level fusion and decision-level fusion approaches are investigated using KELM [2] in this paper, we propose to use human limbs to augment constraints between neighboring human joints. We model a limb as a wide line to represent its shape information. Instead of estimating its length and rotation angle, we calculate as per-pixel likelihood for each human limb by a ConvNet. [3] Video based action recognition is one of the important and challenging problems in computer vision re- search. The several realistic datasets are the HMDB51, UCF50, and UCF101. BoVW is a general pipeline to construct a global representation from local features composed of five steps; (i) feature extraction, (ii) feature pre-processing, (iii) codebook generation, (iv)feature encoding, and (v) pooling and normalization. [4]For efficient video representation on action recognition is shown by dense and achieved state-of-the-art results on a variety of datasets. The performance is corrected by taking the camera motion into account. The feature points between frames using SURF descriptors and dense optical flow are used to

estimate camera motion. With the help of RANSAC homography is estimated. The performance can be significantly improved by removing background trajectories and warping optical flow with a robustly estimated homograph approximating the camera motion. [5] The discriminative features for action recognition Convolutional Neural Networks (ConvNets) are employed into color texture images that are referred to as skeleton optical spectra. The learning of suitable dynamic features and ConvNet architecture from skeleton sequences is possible without training millions of parameters from this kind of spectrum views.

3. DESIGN METHODS

There are some methods used for Human Action Recognition using RNN.

3.1 HUMAN SILHOUETTE EXTRACTION

In this project we extract human body contour from given image. In Frame Videos: To obtain the human body we should take the direct difference between the background and the current frame. Out Frame Videos: To extract the human body from frames of the videos we used the inbuilt Gaussian Mixture Model based Foreground detection method.

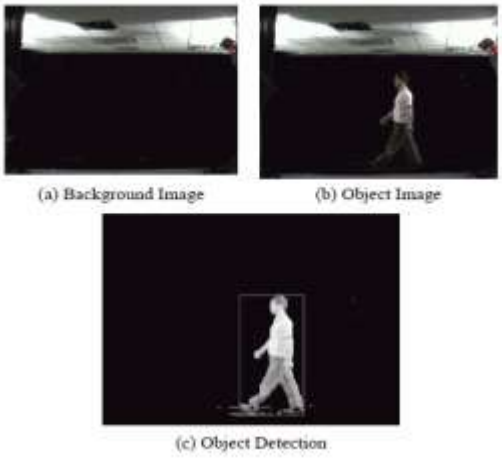

Fig 3.1 Background Subtraction

3.2 HUMAN CONTOUR EXTRACTION

To extract the contour of a detected human body, at thresholding and morphological method, the important approaches in the field of image segmentation are to choose a correct threshold and that is difficult under irregular illumination.

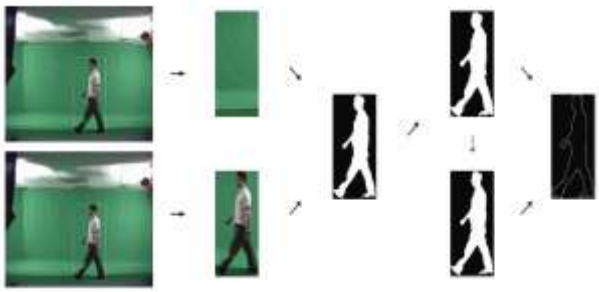

Fig 3.2 Extraction of Human Body Contour

3.3 STAR SKELETONIZATION

The concept of star skeleton is to connect from centroid to gross extremities of a human contour. To find the gross extremities of human contour, the distances from the centroid to each border point are processed in a clockwise or counter-clockwise order. The star skeleton is constructed by connecting the points to the target centroid.

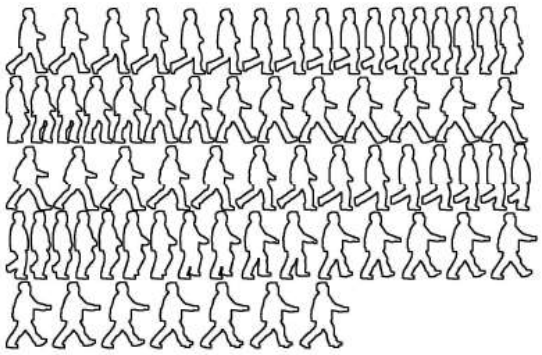

Fig 3.3 a walk action is a series of postures over time

3.4 TRAINING THE MODEL USING RNN

From the class of artificial neural network the connection between nodes form a sequence of directed graph. To learn the temporal dynamics of sequential data that contains the cyclic connections from the neural network architecture. To process sequence of inputs RNN use their internal memory.

4. CONCLUSION

A Recurrent neural Network (RNN) methodology for Human Action Recognition using star skeleton as a representative descriptor of human posture. Star skeleton could be a quick skeletonization technique by connecting from center of mass of target object to contour extremes. To use star skeleton as feature for action recognition, we have a tendency to clearly outline the feature as a five-dimensional vector in star fashion as a result of the top and 4 limbs are typically local extremes of human shape. In our project we

assumed an action is composed of a series of star skeletons over time. Therefore, time-sequential pictures expressing human activity square measure remodeled into a feature vector sequence. Then the feature vector sequence must be transformed into symbol sequence so that RNN can model the action.

5. FUTURE WORK

Human Activity Recognition (HAR) mistreatment smart phones dataset associated an LSTM RNN. Classifying the type of movement amongst Five categories:

- WALKING_UPSTAIRS,
- WALKING_DOWNSTAIRS,
- SITTING,
- STANDING,
- LAYING

6. REFERENCES

[1] 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.

[2] 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

[3] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition," Comput. Vis. Image Understand., vol. 150, pp. 109–125, Sep. 2016.

[4] H. Wang and C. Schmid, "Action recognition with improved trajectories," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 3551–3558.

[5] 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.