Human Action Recognition System Using CNN

Dr. S. S. Lokhande
Associate Professor, Department of E&TC
Sinhgad College of Engineering, Pune, 411041

Patil Snehal Santosh
Department of E&TC
Sinhgad College of Engineering, Pune, 411041

Abstract
Human action recognition from video scene is becoming a more attractive research topic in computer vision due to its several applications like video surveillance, virtual reality, feature extraction, feature vector calculation and intelligent human-computer interactions. However, accurate recognition of actions is still an open task for the researchers due to cluttered backgrounds, occlusions, and viewpoint variations. Many attempts have been made in the last few decades in order to build a robust and effective framework for action recognition and prediction. This paper gives a detailed review of recent and state-of-the-art research advances in the field of human activity classification.

Keywords: *Human Action Recognition, CNN (Convolution Neural Network), Feature Extraction, Feature Vector Calculation*

I. INTRODUCTION

Human action detection in videos are emerging topics in computer vision since understanding human action helps in management, summarization, retrieval and sports analytics of videos. In the past two decades, significant progress has been made with the invention of local invariant features and a bag of features representation. The task is challenging due to variations in action performance, background settings, and inter-personal differences. Based on the skeletons of humans, found in video sequence the pose of the subjects can be estimated in each frame. The grouping of several frames in timely succession then delivers an action.

There are two things to be noted-Action recognition and action localization (M.Muthu et al., 2018). One is 'what action is performed in the video and the other is 'where the action' of interest is taken place.

Moreover, actions are divided into human actions like walking, running, jogging, human-human interactions such as handshaking, kissing, punching and human-object interactions like calling, writing, driving car. Due to environment variations such as moving backgrounds, different viewpoints or occlusions the detection and classification of actions is even more difficult.

For example, in sports applications, it can help to recognize and understand player's poses allowing to make a good decision on players' fouls, especially for football games that require correct decisions in many situations.

Action detection involves locating actions in space and classifiers are used to identify the action class and their Spatial, Temporal locations. The action recognition approaches can be labeled as local or global approaches (NilamNur Amir Sjarif et al./ 2015).

i.   The local approach of video analysis uses local interest points where each interest point has a local descriptor to describe the characteristics of a point. (i.e. local features describe the image patches which is small group of pixels)

ii.   The global approach of video analysis uses the entire imagei.ehave the ability to generalize an entire object with a single vector.

iii.   Scale Invariant Feature Transform (SIFT) (B Jagadeesh et al., 2016) and Space-Time Interest Points (STIP) (B Jagadeesh et al., 2016) Yanshan Li et al., 2017) are the most commonly used local descriptors for videos.

Challenges in Existing System

Action detection is one of the most challenging tasks in video processing that leads the recognition system to poor efficiency. One of the reason is that each actor has its own style of performing an action, that leading to many variations in the subject's movement. Some of the main challenges are listed below:

i. Environment variations such as moving background, occlusions, different viewpoints.

ii. Variation of actors movement, each person has own style of executing an action.

iii. Various types of activity: gestures, actions, interactions, group activities.

iv. Insufficient amount of training videos.

So there is need for the system that can identify the human action in better way. An attempt is made using 3 - D CNN to increase the accuracy in action recognition.

II. LITERATURE SURVEY

(B Jagadeesh et al., 2016) gives the video based human action detection and recognition is addressed and performed on KTH dataset and on real-time videos. At first, hundred frames are extracted from each video sequence and optical flow between the frames is computed. The data extracted using optical flow is converted to binary image. Then Histogram of Oriented Gradient (HOG) descriptor is used to extract feature vector from the binary images. These feature vectors are given as training features to Support Vector Machine (SVM) classifier to prepare a trained model. And finally the actions like walking, jogging, running, boxing, hand waving and handclapping are recognized from video.

(NezihaJaouedi et al.2016) presented a system to analyze human behavior video scene to extract hidden behind gesture, sudden motion and walking speed. Here K Nearest Neighbors approach is used to recognize human behavior through motion analysis.

(Chaitra B H et al.2013) developed algorithm for the human action recognition system, which uses the two-dimensional discrete cosine transform(2D-DCT) for image compression and the self-organizing map(SOM) neural network for recognition purpose, is simulated in MATLAB. 2D-DCT is used to extract image vectors and these vectors become the input to neural network classifier, which uses self-organizing map algorithm to recognize elementary actions from the images

(Mona Moussaet al.2013) presents a fast and simple method for human action recognition. It firstly used interest points using SIFT (scale invariant feature transform) from each frame of the video. Then the popular approach Bag of Video Words is applied with a new normalization technique. This normalization technique remarkably improves the results. Finally a multi class linear Support Vector Machine (SVM) is utilized for classification.

(ZhongxinQuetal.2015) combined discrete HMM model with improved LBG VQ for the action recognition. The input of the discrete HMM must be discrete value.

(Dipankar Das et al. 2014) gives action identification scheme that attempt to evaluate the qualitative characteristics of the human activity. To recognize the activity of human this method initially extracts HOG characteristics of action for certain duration of period to construct an activity model known as the histogram of oriented gradient pattern history (HOGPH). After that the HOGPH characteristics vector are used to categorize the human actions with support vector machine (SVM) classifier.

The stationary information as well as activity information is required to get stout categorization results in human action identification. The detection method is estimated based on the KTH**(Florian Baumann [9])**data set for single activity identification with original training/testing opening and a 5-fold cross verification that enhance the overall accuracies on a normal workstation.

To recognize the complex activities automatically and effectively from video **(Kanchan Gaikwadet al.2010)** proposed Elman Network (EN) and two hybrids of Neural Network (NN) and HMM, i.e. HMM-NN and NN-HMM. There are many types of action recognition algorithms. Some of the classical methods are explained in **(He Chun-Lin1 ,a Pan We et al., 2015)**. Conditional random field (CRF) **(Wang S B et al., 2006)** is a discriminant undirected graph probability learning model, is applied to label and segment ordered data. CRF is essentially a discriminate markov model better adaptive to the real-world condition. CRF is appropriate for dynamic time series modeling, and it can better recognize the simple actions, continuous and complex interaction behaviors.

III. PROPOSED SYSTEM

Identification of human behavior from video has explored the visualization based research much more interesting. It plays important role in image processing, intelligent video surveillance, human-computer interaction and content based video information retrieval. The action can be recognized by extracting feature of movement of targeted object. The system can be used in video surveillance and monitoring, system to analyze the action of the people in terms of security. Figure shows the architecture of the proposed system that will implemented practically.
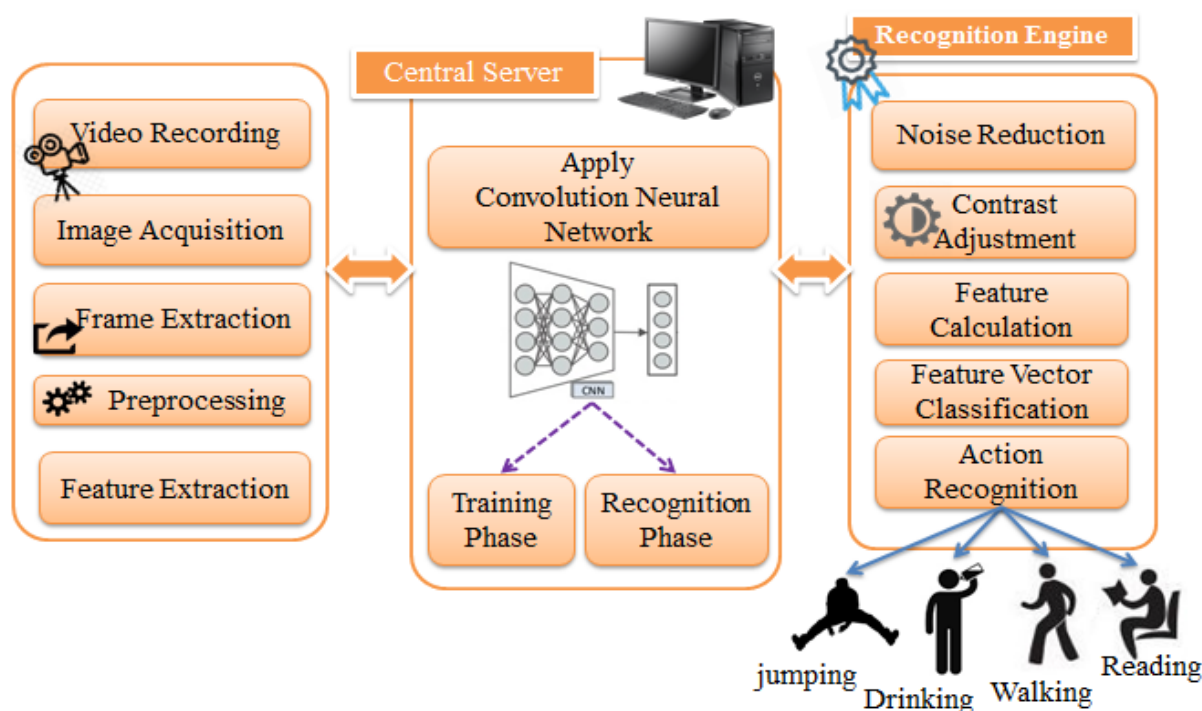
System Architecture

The proposed system helps for recognizing human activities from video sequences with the help of CNN. The system can be used in video surveillance and monitoring, system to analyze the action of the people in terms of security.

i.   Firstly frame extraction on is done from input video.
ii.   By applying preprocessing techniques the noise is removed from the image frame to get the clear fine-tuned image.
iii. The feature extraction is done using CNN from the images.
iv.   Features are calculated for the Actions, Activities, and Events.
v.   Then by using trained model we can classify the action by applying Convolution Neural Network (CNN).

IV. CONCLUSION

Human Activity Recognition (HAR) is popular exploration area in computer vision and image processing area. A large collection of methods are identified in the literature. This paper provides different approaches used by the researchers for recognizing human action. The survey reveals the milestones achieved during the past five or six years in recognising human actions and activities. Here an attempt is made on recognizing the actions by pose detection using 3D- CNN algorithm. By using the output of the trained artificial neural network with a manual pose and action classification the action can be determined.

REFERENCES

[1].   Priyanka Kokate,Surendra Gupta,"Behavior Analysis fromVideos Using Motion based FeatureExtraction",International Research Journal of Engineering and Technology(IRJET),Volume: 06 Issue: 04 | Apr 2019.
[2].   B Jagadeesh ; Chandrashekar M Patil ,"Video based action detection and recognition human using optical flow and SVM classifier", 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).
[3].   M.Muthu Lakshmi, 2Dr.P. Arockia Jansi Rani,"Human Action Recognition",IJEDR2018| Volume6, Issue 2| ISSN: 2321-9939.
[4].   Neziha Jaouedi ; Noureddine Boujnah ; Oumayma Htiwich ; Med Salim Bouhlel, "Human action recognition to human behavior analysis", 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT).
[5].   Chaitra B H, Anupama H S,Cauvery N K,"Human Action Recognition using Image Processing and Artificial Neural Networks",nternational Journal of Computer Applications (0975 –8887)Volume 80 –No.9, October 2013.
[6].   Mona Moussa,Elsayed Hemayed,Heba Elnemr,M. Fayek,"An enhanced method for human action recognition",Journal of Advanced Research • December 2013,DOI: 10.1016/j.jare.2013.11.007 3
[7].   Zhongxin Qu ; Tingshu Lu ; Xiaojiong Liu ; Qianqian Wu ; Mingjiang Wang ,"A new method for human action recognition: Discrete HMM with improved LBG algorithm", 2015 IEEE 9th International Conference on Anti-counterfeiting, Security, and Identification (ASID).

[8]. Dipankar Das, "ACTIVITY RECOGNITION USING HISTOGRAM OF ORIENTED GRADIENT PATTERN HISTORY", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 4, No.4, August 2014.

[9]. Florian Baumann,"Action Recognition with HOG-OF Features".

[10]. Kanchan Gaikwad,"HMM Classifier for Human Activity Recognition",Computer Science & Engineering: An International Journal (CSEIJ), Vol.2, No.4, August 2012.

[11]. He Chun-Lin1 ,a Pan Wei"Study on human Action Recognition Algorithms in videos",2015.

[12]. Wang S B, Quattoni A, Morency L, et al. Hidden conditional random fields for gesture recognition//Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, 2: 1521-1527.

[13]. Yanshan Li,Rongjie Xia,Qinghua Huang,Weixin Xie,"Survey of Spatio-Temporal Interest Point Detection Algorithms in Video"June 2017, DOI: 10.1109/ACCESS.2017.2712789.

[14]. Nilam Nur Amir Sjarif ,Siti Mariyam Shamsuddin,"Human Action Invarianceness for Human Action Recognition ",2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)