Camera Based Action Detection Using Artificial Intelligence

Shraddha Pandit[1], Shraddha Aher[2], Chinmayee Kudtarkar[3], Sayali Patil[4], Prof. Gauri Salunkhe[5]

(Department of Electronics and Telecommunication, Atharva College of Engineering Malad-400095, India
(Department of Electronics and Telecommunication, Atharva College of Engineering Malad-400095, India
(Department of Electronics and Telecommunication, Atharva College of Engineering Malad-400095, India
(Department of Electronics and Telecommunication, Atharva College of Engineering Malad-400095, India
(Department of Electronics And Telecommunication, Atharva College of Engineering Malad-400095, India

Abstract: *Human action recognition, based on video inputs, has become one of the most renowned research areas in the fields of computer vision, and pattern recognition. It has wide variety of applications such as surveillance, robotics, health care, video streaming and human-computer interaction. In our project, we are going to use it for surveillance purpose. Here, we are using the system to monitor real-time events taking place at the same venue. The main objective of our system is to detect different actions using Artificial Intelligence. An immediate action will be taken by the system itself so as to reduce human interference. Particularly, Recurrent Neural Networks (RNNs) are used to represent the track features. The attended features are then processed using other RNN for event detection/classification.*

Keywords: *Real Time Action Detection, Artificial Intelligence, Recurrent Neural Network, Raspberry Pi 3, CCTV Camera*

I. Introduction

Since the invention of computer and machines, their capability to perform various tasks went on growing exponentially. Humans have evolved in such a way that they have acquired the power of computer system in terms of the diversity in their operating domains, increased speed, and reducing size with respect to time. A branch of computer science, named Artificial Intelligence, pursues creating the computer or machines as intelligent as human beings. Artificial Intelligence is a way of transforming computer into a controlled robot, or to make software think intelligently in a manner in which the intelligent humans think.

Artificial Intelligence is accomplished by studying how human brain thinks, and how humans learn, decide and work while trying to solve a problem, and then using the outcomes of this study as a basic of developing intelligent software and systems. Human action recognition, based on video inputs, has become one of the most renowned research areas in the fields of computer vision, and pattern recognition. It has a wide variety of applications such as surveillance, robotics, and health care, video-streaming and human computer interaction.

In our project, we are using it for surveillance. Here, we are using this system to monitor real time events taking place at same venue. The main objective of our system is to detect different actions using AI. An immediate action will be taken by the system itself so as to reduce human interference. Particularly, Recurrent Neural Networks (RNNs) are used to represent the track features. The attended features are then processed using other RNN for event detection/classification.

The system, designed by us, consists of two parts, one is CCTV camera part with the Raspberry Pi 3 module and the other is for Artificial Intelligence software part. The system constantly monitors the actions using the camera. The camera output is constantly processed by the Raspberry Pi 3 module and if any action is detected, the AI system processes the data and the details are stored on the USB storage of the Raspberry Pi. Thus, our system can be used for various purposes, such as theft detection, supermarkets, jewellery shops, where CCTV arrangements are feasible. This paper aims to developed a real time action detection algorithm with high performance based on RNN.

II. Literature Review

In the last five years, different types of deep learning techniques have been applied on single viewpoint datasets. This is not only because single viewpoint human action recognition is the foundation of the research area and provides large scale datasets, but also the framework which has been developed for single viewpoints can be directly extended to multiple viewpoints by generating multiple networks.

Human action detection is currently one of the One of the active research topic in computer vision. Wide range of applications for tracking and behavior recognition studied by many researcher. Automated system for estimating and tracking moving object have received a lot of attention from various industries and academics for its potential application in the field of surveillance and engineering.

In recent times, recognition has become one of the most renowned research areas in various fields such as computer vision and pattern recognition. It has a wide variety of applications in surveillance purpose. Multiple viewpoint datasets contain information from multiple cameras from different directions which naturally avoids the drawback of occlusion and it also captures different view of same gestures from different angles, thereby it provides more information for better performance.

In [1], this system described a rule based algorithm for robust human action detection combined with event recognition using a recurrent convolution neural network (RCNN).

Different architecture, approach, programming language, processor and memory requirement, data based for training/testing images and performance measure of human detection system used in each study of different ANN algorithm each study has its own strengths and limitation. In [4], their system introduced an extension to the RNN encoder-decoder model which aligned the words generated in translation in sequence. In the last two decades, a significant development in human motion recognition and analysis has been observed.

To classify the previous work in the domain, the researchers used criteria like type of models, the dimensions of the tracking space (2-D versus 3-D). Some reviews classify the literature using complexity of the action to be identified [5].

In [3], the recurrent network acquires information obtained from the individual glimpses and combines the information in a coherent manner that preserves spatial information. The glimpse feature vector from the glimpse network is supplied as input to the recurrent network at each time step. The recurrent network consists of two recurrent layers with non-linear function.

In [19], Structural Recurrent Neural Networks (SRNN) is a novel framework for modeling interactions which uses deep architectures. It can be used for merging the capabilities of high-level spatial-temporal graphs and sequence-learning features of Recurrent Neural Networks (RNNs).

There has been a fair amount of research on activity analysis, and recent surveys can be found [6], [7]. Estimating a human pose using a predefined model in each frame is the most common technique for generic human model recovery. The model is driven by an attempt to minimize the cost function between the collection of parts arranged in a deformable configuration and human contours [7], [8].

III. Proposed System

This system, known as Camera Based Action Detection using AI, is designed for detection of various human actions. Using this system, it is possible to monitor real time events in a room or any such small area. Our system is used in many places like shopping malls, jewellery shops, super markets also it can be used for theft detection where CCTV arrangements are feasible.

CCTV stands for Closed Circuit Television. It is also known as video surveillance and can be used as video cameras to transmit a signal to a remote place. There are different types of CCTV cameras like Bullet camera, C-mount camera, Dome camera, etc. In our system, we are using Dome CCTV camera for real time monitoring or detecting actions because it is IR night vision camera. Also, it is easy to install.

Recurrent neural networks (RNNs) are type of neural networks that were designed to learn from sequence data, such as sequences of observations over time, or a sequence of words in a sentence. RNNs have proved to be effective in tackling challenging sequence prediction problems [4][10].

In [19], SRNN has shown its impact in spatial-temporal applications such as human motion modeling, human activity detection, anticipating driver maneuvering, 3D action recognition etc.

The image/video captured by the CCTV camera will be stored in the 128 GB memory card. The processing of these stored images will be done by the Raspberry Pi module, which will further detect whether the intended action is captured or not. In our system, RNNs are being used in order to identify whether the person entering the monitored place is an authorized one or not. If an unauthorized person enters the place, the alarm will be indicated to the authorized personnel. If the captured image matches with the image stored in the database, then it will be detected on the monitor by changing the colour of the monitoring screen to red. Thus, human action will be captured. Also, it is possible to live stream the human actions.
.

IV. Requirement Specifications

Computer software's need both hardware and software resources to be present on a computer in order for its optimum utilization. These prerequisites are known as system requirements. Also, these are often used as a guideline for the optimized performance. Generally, software defines two sets of system requirements:

minimum and recommended. System requirements tend to increase over time with increasing demand for higher processing power and resources in newer versions of software.

A. Hardware Requirements: It is the most common set of requirements which are defined by any OS or application. These are basically the physical computer resources. The following subsections discuss the various aspects of hardware requirements:

• Raspberry Pi 3
• USB storage 128 GB
• Dome CCTV camera

B. Software Requirements: These are the requirements and prerequisites that need to be installed on a computer to obtain optimal functioning of an application. Generally, these requirements are not included in the software installation package and need to be installed separately before the software is installed. Requirement Specification: Operating System: Windows OS (XP, 7 and further versions), Mac/ Linux.

V. Implementation
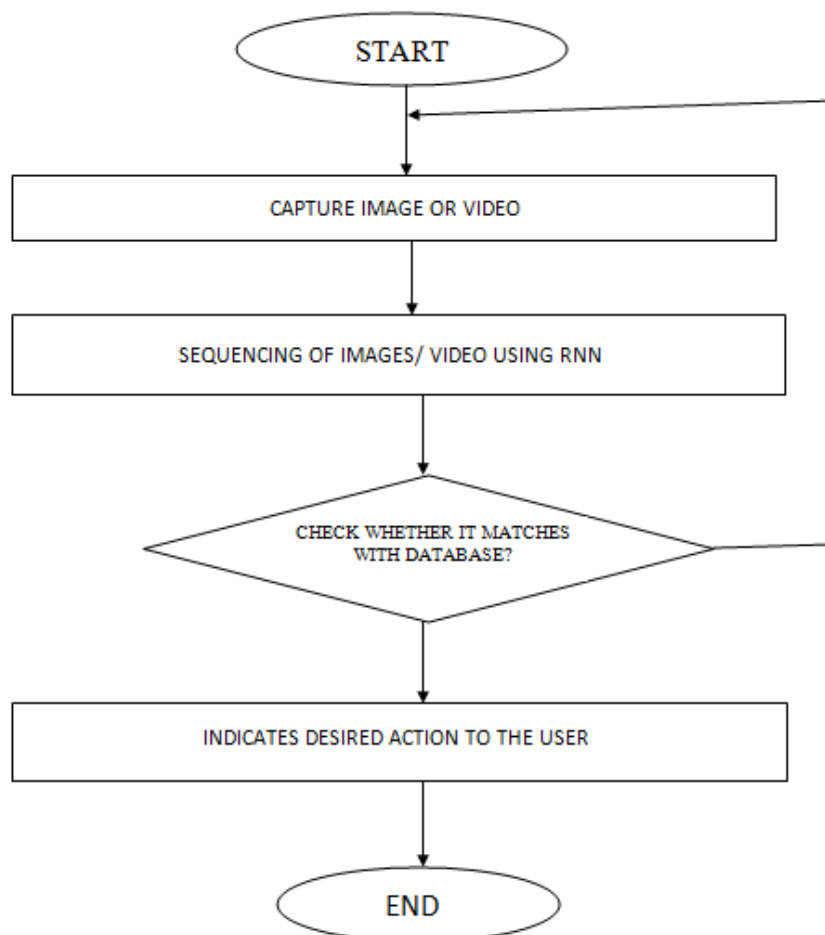
VI. Conclusion

• Artificial Intelligence has been widely used in speech recognition, language processing and recommendation system.
• It is possible to design a real time, adaptive and high performing recognition system.
• Most artificial intelligence system have the ability to learn, which allows people to improve their performance over time.
• The operation of AI is based on accessing huge amounts of information, processing it, analyzing it and then executing tasks to solve certain problems based on its operation algorithms.

References

[1]. M. Turan, Y. Almalioglu, H. Araujo, E. Konukoglu, and M. Sitti, "Deep endovo: A recurrent convolutional neural network (rcnn) based visual odometry approach for endoscopiccapsulerobots,"

[2]. Neurocomputing, vol. 275, pp. 1861–1870, 2018.

[3]. S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. In Advances in neural information processing systems, pages 561–568, 2002. 5, 6.

[4]. J. Ba, V. Mnih, and K. Kavukcuoglu. Multiple object recognition with visual attention. arXiv preprint arXiv:1412.7755, 2014. 2

[5]. D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473, 2014. 2, 4, 5

[6]. Aggarwal, J. K., &Cai, Q. (1997, June). Human motion analysis: A review. Proceedings of theNonrigid and Articulated Motion Workshop '97 (pp. 90-102). IEEE. doi:10.1109/NAMW.1997.609859

[7]. R. Poppe, "A survey on vision-based human action recognition," Image Vis. Comput., vol. 28, no. 6, pp. 976–990, 2010.

[8]. B. Sapp, D. Weiss, and B. Taskar, "Parsing human motion with stretchable models," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011, pp. 1281–1288.

[9]. Cherian, J. Mairal, and K. Alahari, "Mixing body-part sequences for human pose estimation," in in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2014, pp. 2361–2368.

[10]. Cross-Layer Design of Coded multicast for wireless random access network by Ketan Rajawat, Nikolaos Gatsis, Seung-Jun Kim, And Georgios B. Giannakis.

[11]. Deep Learning Models for Human Activity Recognition by Jason Brownlee.

[12]. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition By Yong Du, Wie Wang, Liang Wang National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences.

[13]. Gesture Recognition using Artificial Neural Network By Khushboo Arora, Shrutika Suri, Divya Arora, Vaishali Pandey A.P, CSE, MRIU, India.

[14]. Detecting events and key actors in multi-person videos By Vignesh Ramanathan, Jonathan Huang, Sami Abu-El-Haija, Alexander Gorban, Kevin Murphy and Li Fei-Fei Stanford University, Google.

[15]. Artificial Intelligence and International security By Michael Horowitz, Paul Scharre, Gregory C. Allen, Kara Frederick, Anthony Cho and Edoardo Saravalle.

[16]. Human Motion Detection and Tracking for Real-Time Security System by Pavithra S., Mahanthesh U., Stafford Michahial, Dr. M. Shivakumar.

[17]. Advances in Human Action Recognition: A Survey By Guangchun Cheng, Yiwen Wan, Abdullah N. Saudagar, Kamesh Namuduri, Bill P. Buckles.

[18]. Whitepaper on 'Will AI Change the game for cyber security in 2018?' by Rajat Mohanty, CEO, Paladion.

[19]. A Review of Human Activity Recognition Methods by Michalis Vrigkas, Christophoros Nikou, Ioannis A. Kakadiaris.

[20]. Human Weapon-Activity Recognition in Surveillance Videos Using Structural RNN by Praneeth Kumar Susarla, Utkarsh Agarwal, Dinesh Babu Jayagopi from University of Oulu and IIIT Bangalore respectively.