HUMAN ACTION DETECTION AND RECOGNITION BY USING DEEP CONVOLUTIONAL NEURAL NETWORK

1 Assistant Professor, 2,3,4 UG Scholar,
1,2,3,4 Department of ECE, Francis Xavier Engineering College, Tirunelveli.
pradheeptrajan@gmail.com

ABSTRACT
This paper explains a deep learning method to detect and recognize human actions. In this, a camera is used to monitor the human actions. A deep convolutional neural network is used for the image or video obtained from the camera. At first segmentation is performed followed by several processing techniques. After this the human actions are detected and recognized by using deep convolutional neural network. Finally, the recognized action is send to a particular person by using ARDUINO. It is developed mainly for two applications: 1. neural network for elder people or active and assisted living system for smart home, 2. Smart gesture identification. This project is implemented using MATLAB simulation and ARDUINO.

Keywords—Neural network, MATLAB, ARDUINO, neural network, Smart gesture.

I. INTRODUCTION
Human action recognition plays an important role in interaction between human and computer. It have been already used widely in electronic markets. Human action recognition has so many applications. One of the common applications is, they have been commonly used to monitor the elder people at the home. Due to the busy lifestyle, they are needed a lot for elder people monitoring. This paper uses a camera to monitor the action of the human. The camera do not give any facial information in order to maintain the privacy concern. The video taken from the camera is fed as the input. After segmentation, the segmented RGB image is converted into gray scale image to make the process much easier. Finally, the gray scale image is compared with the deep convolutional neural network algorithm

II. EXISTING SYSTEM
Actually a bulk of works are available related to the human action detection. Each one have their own limitations. In [15] a depth camera and inertial sensor is used to detect the human actions automatically. In [10] a inertial signal, skeleton joint position and depth camera is used to detect the human actions more accurately. In [22] both CNN and LSTM are used to detect large amount of human actions. Mostly many of the related uses skeleton joint position in order to accurate result. Some uses additional sensors to make the process easier.

For the smart TV application, the continuous dataset for the smart TV hand gestures in [13] is used here. The actions or gestures of interest in this dataset consist of ‘waving a hand’, ‘flip to left’, ‘flip to right’, ‘counter clockwise rotation’ and ‘clock-wise rotation’. For these gestures, the inertial sensor was worn on the wrist. This dataset contains 5 continuous gesture streams each containing the above 5 gestures from 12 subjects.
Fig. 1 Block diagram of the detection and recognition fusion system

A. SEGMENTATION

For segmentation of the transition movements in the depth camera path, the centroid (cx; cy) of the background subtracted depth images is obtained as follows:

\[ c_x = \frac{\sum_{i=1}^{N} x_i m_i}{\sum_{i=1}^{N} m_i}, \quad c_y = \frac{\sum_{i=1}^{N} y_i m_i}{\sum_{i=1}^{N} m_i} \]

B. CNN+LSTM ARCHITECTURE FOR INERTIAL SIGNALS

The first stream uses CNN and LSTM layers and the second stream directly uses the handcrafted inertial features. The input to the first stream is 8 time-series signals corresponding to the 3-axis acceleration signals, the 3-axis angular velocity signals, the overall acceleration signal, and the overall angular velocity signal. The overall acceleration is obtained using the acceleration at each frame t of a sequence as follows.

III. PROPOSED SYSTEM

Initially a video with the action of a single human is taken. When they are fed as the input several image processing steps are performed to make the process easier.

A. PREPROCESSING

At first, pre-processing with median filter occurs. Because the input data may have some difference from the trained data in both horizontal and vertical plane. In order to remove the difference between them, the median filter is used. The differenced data between them is a waste, so they are totally removed.

B. MORPHOLOGICAL PROCESS

Morphological process is nothing but the ordering of the pixel. It is more effective when they are applied to the gray scale image.

C. RGB TO GRAY IMAGE

In this, the RGB input is converted into gray scale input. The RGB input is converted into the gray scale input to perform the morphological process effectively.

D. HARDWARE DESCRIPTION

Arduino is an open-source platform used for building electronics projects. Arduino consists of both a physical programmable circuit board (often referred to as a microcontroller) and a piece of software, or IDE (Integrated Development Environment) that runs on your computer, used to write and upload computer code to the physical board. Finally, Arduino provides a standard form factor that breaks out the functions of the micro-controller into a more accessible package.

IV. RESULT AND DISCUSSION

Fig. 3. Gray Scale Input
The above figure explains the gray scaled input is a video. In order to make the action detection easier, the video is segmented i.e. they are converted into several frames or pictures. The CNN (Convolutional Neural Network) is an algorithm. The frames are compared with the trained data. By this the human actions is detected. The actions are detected and it will be displayed at the output screen.

Next, by using GSM and ARDUINO the detected action is send to the specified person. By this GSM module we can send to so many person by the already fed number in the MATLAB program. In this, the ARDUINO is just a intermediate between the software and hardware setup.

The development of deep learning method, for human action detection and recognition was tested with several data. In this the trained deep CNN plays a major role. The coding for both action detection and their sending process is written in the MATLAB.

IV. CONCLUSION

In this paper, a deep learning method for the detection and recognition of human action is developed. In this, a camera is used to monitor the human actions. A deep convolutional neural network is used for the image or video obtained from the camera. At first segmentation is performed followed by several processing techniques. After this the human actions are detected and recognized by using deep convolutional neural network. Finally the recognized action is send to a particular person by using ARDUINO. It is developed mainly for two applications: 1. neural network for elder people or active and assisted living system for smart home, 2. Smart gesture identification. This project is implemented using MATLAB simulation and ARDUINO. They are applicable for the real time processes. This system uses a camera, GSM, ARDUINO and a deep CNN algorithm. The result explains the development of human action detection and recognition process.

REFERENCES


Dr. N. Muthukumaran, Dr. R. Joshua Samuel Raj, Manjula. R, Pavithra. L, Nagalakshmi. T, 'Automatic Identification and Management in Parking Lot through IOT', International Journal of


