Gesture Recognition using CNN and RNN

Rajalakshmi J, Kumar P

Abstract- Gesture Recognition is a major area in Human-Computer Interaction (HCI). HCI allows computers to capture and interpret human gestures as commands. A real-time Hand Gesture Recognition System is implemented and is used for operating electronic appliances. This system is implemented using the deep learning models such as the Convolution Neural Network (CNN) and the Recurrent Neural Network (RNN). The combined model will effectively recognize both static and dynamic hand gestures. Also the model accuracy while using VGG16 pre-trained CNN model is investigated.

Keywords: Convolution Neural Network (CNN), Human-Computer Interaction (HCI), Recurrent Neural Network (RNN).

I. INTRODUCTION

Human-computer interaction is how humans interact with computers. HCI focuses on creating systems that are easy to use by users. HCI aims to make the computers more user-friendly and more reactive to the user needs. It has its application in various fields such as gaming, education, medical applications, emergency planning, and response applications. Hand gesture recognition has considered being an excellent means of HCI. Gestures are nothing but the nonverbal means of communication which includes bodily actions to convey some messages, these messages can be used as commands to the system. A gesture can be classified into two types' static gestures and dynamic gestures. Static gestures are postures that do not consider movements into account, e.g. thumbs up. Dynamic gestures consider the angles between fingers during certain start and end times, e.g. drawing letters in the air. Machine learning is used to solve various real-time problems. It is commonly used for classification, detection, recognition, and predictions problems. Machine Learning is used to automate processes using data. A model is produced and is capable of returning an output, this output may be the new data or some predictions with the known data. Deep learning is a class of machine learning that uses more number of layers to extract higher-level features from the input. The word deep refers to the number of layers through which the data is transformed.
Hand gesture recognition where initially glove based control. This system has various applications not only in human-computer interaction but also in gaming, sign language recognition and so on. Due to the development in the glove based type of systems offers system the ability of vision-based recognition without the need of any sensors. The Vision-Based gesture recognition system uses a camera for capturing the gesture in the image or video format. This approach is more user-friendly but is difficult to implement and it does not require any additional devices for gesture analysing.

Revised Manuscript Received on June 22, 2020.

Rajalakshmi J, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, India. E-mail: j.rajalakshmi96@gmail.com

Dr. P. Kumar, Professor, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, India.

The vision-based approach considers properties such as texture and colour and the limitations of this approach are lighting, location, image noise. In glove or sensor-based approaches, the user needs to wear a glove with is associated with a sensor to determine the hand gesture. Similarly, a bodysuit is required if in case of body gesture recognition. This approach requires multiple sensors and cables to connect those devices. The cost of the sensors and gloves are expensive, this will be the major drawback of this approach. The Depth based approach is another robust approach in gesture recognition. In this approach, the 3D geometric information depth camera is used. This method directly reflects the depth image with the coloured image. The illumination, shadow, and colour of the input do not affect this method. The latest depth-based gesture recognition uses a Kinect device, which is used to find the depth of the object. The limitations of this method are a Kinect camera was more expensive.

II. LITERATURE SURVEY

There are various machine learning algorithms were used for the recognition of hand gestures and few of them are discussed as follows [1] proposes a real-time hand gesture recognition system with four modules such as Data acquisition, Pre-processing, Feature extraction, Gesture recognition, and the Real-time hand detection is done using Histogram of Oriented Gradients (HOG) feature in MATLAB and the k-Nearest Neighbour (KNN) algorithm is used to classify the input images. A technique which commands computer using six static and eight dynamic hand gestures is proposed in [3] with three main steps hand shape recognition, tracing of detected hand (if dynamic), and converting the data into the required command, the system uses VGG16 a CNN architecture is used as the pre-trained model to recognize the gestures. [4] Introduces an effective Hand Gesture Recognition system where the feature extraction is done using Deep Convolutional Neural Network (DCNN) and the classification is done using a Multi-class Support Vector Machine (MCSVM). [5]realizes the segmentation of hand gestures by establishing the skin colour model, the Hand gesture segmentation is done using AdaBoost classifier based on Haar and the CamShift algorithm for hand gesture tracking and the hand gesture area is classified using a convolution neural network. The framework for hand detection and gesture recognition is proposed in [16], the raw data coming from the Kinect is used to get the depth information of an image and an algorithm is proposed to identify each point of a closed contour identified within a given depth interval. Starting from the regained contour points, the centre of the palm is identified. The fingertips are localized by employing the k-curvature algorithm.

III. PROPOSED SYSTEM

A Gesture Recognition System is developed for controlling electronic appliances to provide remote access to electronic appliances without the use of switchboards. The system uses a deep neural network which is the combination of Convolution Neural Network (CNN) and Recurrent Neural Network (RNN). Combining CNN and RNN will enhance the ability to recognize different actions at a varied time, so we can also recognize dynamic gestures. The system comes under the vision-based gesture recognition method. It uses a camera to capture the input hand gesture. The feature extraction and gesture recognition are done in the RCNN model. The system uses VGG16 architecture for feature extraction and gesture recognition. The last 3 layers of the VGG model are replaced with the Recurrent Neural Network which is used to identify the dynamic gestures. The architecture diagram of the Gesture Recognition System is shown in Fig 1.

Initially, the RCNN model is developed by replacing the last 3 layers of the Convolution Neural Network by the Recurrent Neural Network. The proposed system uses a VGG16 architecture which is a pre-trained convolution neural network. The model is trained with the training datasets which consists of various hand gestures at varied lighting and background conditions and tested with a set of new data to determine how well the model works for new

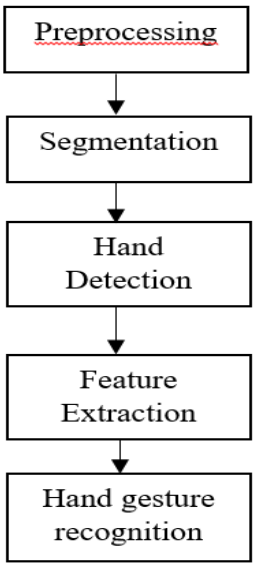

Fig 1: Gesture Recognition System

data. The real-time hand gesture is captured using a camera and is subjected to the pre-processing phase where the background is eliminated also scaling and normalization are carried out. The pre-processed image is next provided to the RCNN network to extract the features from the input gesture. The extracted features are compared with the pre-trained datasets to recognize the gesture. Each gesture is interpreted with a command, here the command will be ON or OFF of the electronic appliances.

This Gesture Recognition system is implemented using four modules namely Dataset collection and Pre-processing, Training and Testing, Hand Detection, Feature extraction and Gesture Recognition.

A. Dataset Collection and Pre-processing

The dataset consisting of various hand gesture images at different conditions such as lightning, varied backgrounds, dimensions and so on are collected and are subjected to pre-processing. The pre-processing of the image consists of the conversion of RGB to a greyscale image, background elimination, scaling.

B. Training and Testing

After pre-processing the dataset, the RCNN model is implemented to recognise the gestures and to interpret the gestures with a command. The model is trained using a training dataset so that the model learns the features of the hand gestures. After training a model, a new dataset is provided for testing. The testing phase helps to check the accuracy of the model.

C. Hand Detection

For a real-time hand gesture, edges of the hand are detected to extract the features. The hand is detected based on the number of fingers in the gesture and orientation of a gesture.

D. Feature extraction and Gesture Recognition

From the detected hand the features are extracted from the gesture and these features are compared with the existing datasets to recognise the gesture. Each recognised gesture is associated with a command that will perform on the electronic appliances.

IV. EXPERIMENTAL RESULTS

Gesture Recognition System can be implemented using various techniques and some of them are listed under the literature survey. One such technique is using Convolution Neural Network for recognising the gesture. This method is further investigated by implementing the system using VGG16, a pre-trained CNN network. The implementation is carried out by pre-processing, feature extraction and gesture recognition. The input data is pre-processed and subjected to the VGG16 architecture where it uses a hierarchical approach for classification and recognition process. It uses a selective search algorithm to extract the features and recognise the gesture. The graphical representation of epoch and loss is shown in

Fig 2: Epoch vs Loss

Fig 2. Initially, the loss is more as epoch increases the loss reduces.

The graphical representation of epoch and accuracy is shown in Fig 3.

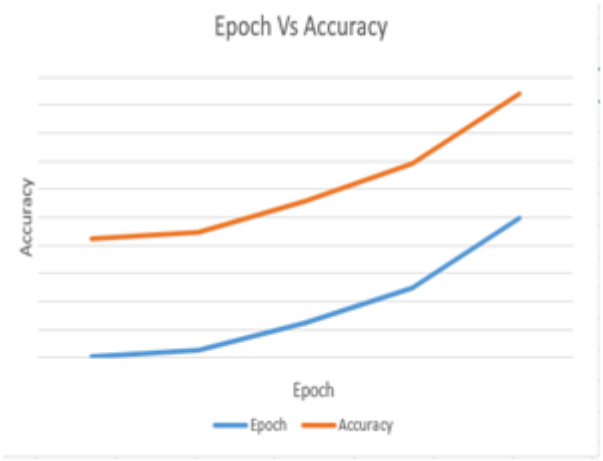

Fig 3: Epoch vs Accuracy

From the graph, it is clear that as epoch increases the accuracy of the model also increases. The average accuracy of the model using VGG16 is found to be 88%.

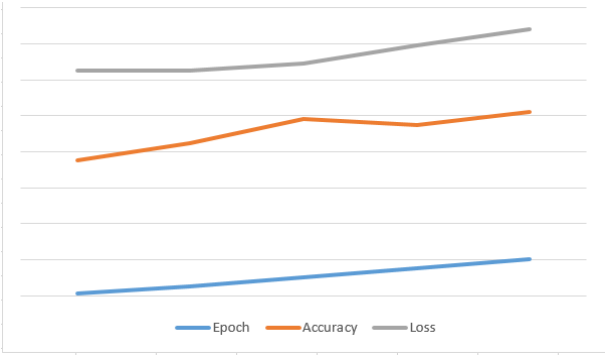

Fig 4: Accuracy vs Loss

The existing model is compared with the proposed work, a combination of CNN with RNN. The last three layers of VGG16 architecture is replaced by RNN and by using a softmax layer to obtain the output. This is used to identify the temporal patters so that it effectively recognise dynamic gestures as well as the static gestures. The proposed model shown in fig 4, obtains more accuracy with minimum number of epoch when compared with the existing model.

V. CONCLUSION

Thus the Gesture Recognition System for controlling electronic appliances is implemented using the Deep learning model by combining the Convolution Neural Network and Recurrent Neural Network. The CNN model will extract the efficient features from the hand gestures and the RNN will identify the temporal patters to recognize the dynamic gestures. In future, more devices can be controlled by including various gestures and will improves the quality of gesture image.

REFERENCE

1. Sanmukh Kaur, Anuranjana (2018), "Electronic Device Control Using Hand Gesture Recognition System for Differently Abled", 8th International Conference on Cloud Computing, Data Science & Engineering.
2. Youngwook Kim And Brian Toomajian (2016), "Hand Gesture Recognition Using Micro-Doppler Signatures With Convolutional Neural Network", IEEE Access Vol. 4.
3. Soeb Hussain, Rupal Saxena, Xie Han, Jameel Ahmed Khan, Hyunchul Shin (2017), "Hand Gesture Recognition Using Deep Learning", International SoC Design Conference.
4. Md Rashedul Islam, Rasel Ahmed Bhuiyan, Ummey Kulsum Mitu, Jungpil Shin (2018), "Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language", 4th International Conference on Frontiers of Signal Processing.
5. Jing-Hao Sun, Shu-Bin Zhang (2018), "Research on the Hand Gesture Recognition Based on Deep Learning", 12th International Symposium on Antennas, Propagation and EM Theory.
6. Anush Ananthakumar (2018), "Efficient Face and Gesture Recognition for Time Sensitive Application", IEEE Southwest Symposium on Image Analysis and Interpretation.
7. Saransh Sharma, Samyak Jain, Khushboo (2019), "A Static Hand Gesture and Face Recognition System for Blind People", 6th International Conference on Signal Processing and Integrated Networks.
8. Mohd. Baqir Khan, Kavya Mishra, Mohammed Abdul Qadeer (2017), "Gesture Recognition using Open-CV", 7th International Conference on Communication Systems and Network Technologies.
9. Arathi P.N, S.Arthika, S.Ponmithra, K.Srinivasan, V.Rukkumani (2017), "Gesture Based Home Automation System", 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software.
10. Kenneth Lai and Svetlana N. Yanushkevich (2018), "CNN+RNN Depth and Skeleton based Dynamic Hand Gesture Recognition", 24th International Conference on Pattern Recognition.
11. P Raghu Veera Chowdary, M Nagendra Babu, Thadigotla Venkata Subbareddy, Bommepalli Madhava Reddy, V Elamaran (2014), "Image Processing Algorithms for Gesture Recognition using MATLAB", IEEE International Conference on Advanced Communication Control and Computing Technologies.
12. Chih-Hung Wu and Chang Hong Lin (2013), "Depth-based hand gesture recognition for home appliance control", IEEE 17th International Symposium on Consumer Electronics.
13. Ransalu Senanayake, Sisil Kumarawadu (2012), "A Robust Vision-based Hand Gesture Recognition System for Appliance Control in Smart Homes", IEEE International Conference on Signal Processing, Communication and Computing.
14. Archana S. Ghotkar, Rucha Khatal, Sanjana Khupase, Surbhi Asati &Mithila Hadap (2012), "Hand Gesture Recognition for Indian Sign Language", International Conference on Computer Communication and Informatics.
15. Lan Tiantian, Shen Jinyuan, Liu Runjie, Guo Yingying (2015), "Hand Gesture Recognition Based on Improved Histograms of Oriented Gradients", The 27th Chinese Control and Decision Conference.
16. Guillaume Plouffe and Ana-Maria Cretu (2016), "Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping", IEEE Transactions on Instrumentation and Measurement, Vol. 65, Issue: 2.
17. Dong-Woo Lee, Jeong-Mook Lim, John Sunwoo, Il-Yeon Cho and Cheol-Hoon Lee (2009), "Actual Remote Control: A Universal Remote Control using Hand Motions on a Virtual Menu", IEEE Transactions on Consumer Electronics, Vol. 55, Issue: 3.
18. Sruthy Skaria, Student, Akram Al-Hourani, Margaret Lech, and Robin J. Evans (2019), "Hand-Gesture Recognition Using Two-Antenna Doppler Radar With Deep Convolutional Neural Networks", IEEE Sensors Journal, Vol. 19, Issue: 8.
19. Guangming Zhu, Liang Zhang, Peiyi Shen, Juan Song, Syed Afaq Ali Shah, and Mohammed Bennamoun (2019), "Continuous Gesture Segmentation and Recognition Using 3DCNN and Convolutional LSTM", IEEE Transaction on Multimedia, Vol. 21, NO. 4.
20. Bin Xie1, Xiaoyu He1, Yi Li1 (2018), "RGB-D static gesture recognition based on convolutional neural network", The Journal of Engineering, Vol. 2018, Issue: 16.

AUTHORS PROFILE

Rajalakshmi J, pursuing her final year M.E. Computer Science and Engineering at Rajalakshmi Engineering College, Chennai. She has published few papers in the reputed Journals. She is currently working in the domain of deep learning.
j.rajalakshmi96@gmail.com

Dr. P. Kumar, Professor, Computer Science and Engineering has been with Rajalakshmi Engineering College, Chennai since June 1998.He has 23 years of teaching and published more than 30 papers in the referred National and International Journals. He has guided many UG and PG projects and one of his project have won Best Project in IBM TGMC. He also published a book on Computer Programming.