CONTROLLING MEDIA OPERATIONS ON AN AUTOMATED LOCAL MACHINE USING COMPUTER VISION

Abhinav Rathore[1], Aditya Rampure R[2], Anwesh Harsh[3], Nair Ashutosh Kamlesh[4]
Dr. Seemanthi K

[1] *BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*
[2] *BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*
[3] *BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*
[4] *BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*
[5]*Associate Professor, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India*

ABSTRACT

In everyday life, people speak with one another and utilize an expansive scope of motions during the time spent connection.Aside from relational correspondence, numerous hours are gone through in the association with electronic gadgets. Hand signal acknowledgment is an arduous assignment to tackle in recordings. In this paper, we use a residual attention network that has been trained from start to finish to recognise hand motions. Given the numerous attention blocks stacked on top of each other, a network is fabricated which makes various highlights at every consideration block. Our consideration based residual network (ResATN) can be fabricated and stretched out to extremely profound layers. Utilizing this network, a broad examination is performed on different networks dependent on three freely accessible datasets. We evaluated our network using one of the finest recognition dataset(Jester) and achieved competitive performances by training the network from scratch. This hand signal recognition system will not only replace the use of a mouse to monitor the media player, but will also include various gesture commands that will be useful in controlling the programme.

Keyword: - *Hand Gestures, ResNet, Jester, CNN, Computer Vision.*

---

1. INTRODUCTION

In discussions with others, we utilize various kinds of signals. In such conversations, non-verbal correspondence (NVC) is a significant part since it could convey up to 55% of the general correspondence. Hand motions structure a piece of our discussion and are critical to completely comprehend the subject being talked about.
It is exceptionally commendable that the connection with the frameworks doesn't basically contrast from the regular association occurring between various clients. Perceptual User Interfaces (PUI) is the premise where they are keen on broadening Human Computer Interaction (HCI) to utilize all modalities of human discernment. Early improvement of PUI, it utilizes vision-based interfaces which perform online hand signal recognition. The best

instruments like mice, joysticks and consoles are able for HCI, as they have been altogether confirmed. In any case, visually impaired individuals can't get to these hand signals and thus will be unable to handily follow the discussion. We use an unusual type of network, such as a residual attention network, which is a start to finish teachable network, to describe the hand signals provided in the video outlines, inspired by the accomplishment of consideration networks and recent developments in deep residual networks.. The purpose behind utilizing residual networks is to have a critical impact on expanding the network profundity. This gives the benefit of deeper networks which are simpler to prepare with an expansion in accuracy. SGD has been used in multi-layer convolutional networks for preparing features but not for preparing the classifier when it comes to object recognition. A significant test advance is unpredictability and strength connected with the investigation and assessment for acknowledgment of motions.

2. SURVEY PAPER

A recent research paper proposed by Javaan Chahl, Munir Oudah and Ali-Al Naji used a camera vision-based sensor, which proved to be more popular, suitable, and applicable because it allows humans and machines to communicate without touching. Several problems, such as background light variance, occlusion effects, complex background patterns, and time processing, were pitted against resolution and rate frame. Color-based identification using the Glove Marker and skin colour, appearance-based recognition, depth-based detection, and 3-D model-based detection can all help solve these obstacles.

Ram Pratap Sharma introduced a paper that had a quick, straightforward and compelling gesture recognition algorithm for robot application that perceives a limited range of motions in a natural way To build a user-friendly environment for human-computer interaction, the author used multi-stream Hidden Markov Models (HMMs) with EMG sensors and a 3-D accelerometer (ACC). The difficulties were that it was only appropriate for simple finite movements against a simple context, not complex gestures. Different segmentation techniques, as well as the Gaussian Model Classifier and other classification techniques, were used to solve these problems.

A paper published by Zhi-hua Cheng, Jung-Tao Kim, Jianning Liang, Jing Zhang and Yo-Bu Yuan proposed the method of finger segmentation. This was accomplished using a complicated set of classifiers such as CRF (Conditional Random Field) and SVM (Support Vector Machines). Some previous attempts required the user to wear a data glove to acquire the gesture showing hand. But this method proved expensive due to the application of a highly expensive data glove. In real life, the authors used a TOF camera , that is , a Kinect sensor to capture and detect the depth of the environment and a tape across the arm to detect the hand signals. This proved to be highly efficient in real time applications.

David J Rios-Soria, Satu E Schaeffer, and Sara E Garza-Villareal presented work on electronic device interaction using Stergiopoulou et al.[SP 09], a self-growing and self-organized neural network for hand gesture recognition. Hirobe et al.[HNW 09] created an interface for gesture control and can be used for recognition of virtual objects. The robotic control mechanism for hand gesture detection was established by Malima et al[CMO006]. The proposed algorithm and machine concepts perform hand gesture recognition by utilising computer vision techniques and can recognize six different gestures in real time.

M.I.N.P Munasinghe introduced an examination paper that clarifies how PC vision-based procedures just as feed-forward neural network based arrangement strategies are utilized to build up a dynamic hand recognition framework. The creator has actualized this methodology in Python utilizing the openCV library created by Intel. This method is a step-by-step approach starting with background subtraction, applying blur to image and applying binary thresholding . Using a neural network, MHIs are continuously generated to classify gestures. Changing of light conditions are additionally included as a boundary to the analysis to distinguish the effect of progress of light Changing of light conditions are additionally included as a boundary to the analysis to distinguish the effect of progress of light.

3. METHODOLOGY

It consists of simple steps where the first step is “downloading a dataset” because we require data to train our model, this is the computer vision field so we require data in the form of pictures. Secondly, we have to prepare our data using a “pandas” library to create dataframes. The third step is “training of model”, in this we’ll see different networks like “CNN and ResNet” on the deep learning model. After training, the trained model will be ready. So this trained model will be used in the classification model, the output of this classification file is the result. After that we’ll use the results in controlling our “Windows Media Player” with the help of windows media player API or keyboard API.

3.1 DATASET

“Jester Dataset” is a huge dataset and it is available for free for learning purposes. It has 27 classes with 148,902 GB images, from which 118,562 for preparation, 14,787 for validation, and 14,743 for research. Using the jester dataset, we'll compare the results of CNN and ResNet-101.

3.2 FLOW DIAGRAM

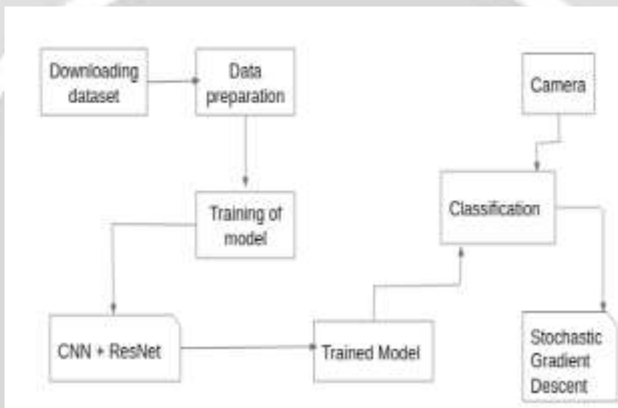

Fig -1: Architecture for the gesture recognition

3.3 CONVOLUTION NEURAL NETWORKS (CNN)

The convolutional layer, rectified linear unit layer, pooling layer, and a fully linked normalised layer are all hidden layers in CNN. To detect patterns, the convolutional layer employs a matrix filter and convolutional operations. The ReLu activation function is applied on the convolutional layer to get the amended feature map. After that the pooling layer uses multiple filters to detect edges, corners, etc and the completely connected normalized layer is our output layer.

3.4 RESNET

The spatial material as well as the temporal relationship between the various frames of a video are captured by CNN. In ResNet and other networks used for comparison, we use CNN. Short connections are direct connections between two non-consecutive layers in the Residual network. Since residual connections make it easier to optimise deeper networks,As a result, accuracy will improve. Our network has many focus blocks and is a residual network. These attention blocks are made up of two parts: a trunk and a trunk and musk base. The trunk layer is made up of residual units, but any other model units may be used in its place. The soft attention differentiable function in the musk layer generates a Mask M(x) of the same dimension as T(x) created by the trunk layer. The attention block's production O can be defined as follows:

$O_{i,c,f} = M_{i,c,f} * T_{i,c,f}$

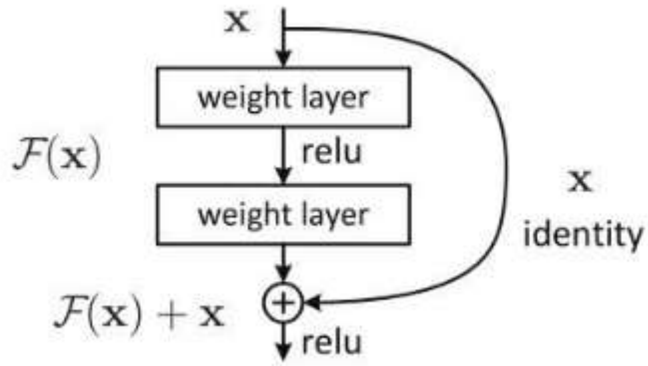

Fig 2: A Residual Block

3.5 STOCHASTIC GRADIENT DESCENT :

It is the optimization algorithm, basically it helps model during training to get itself better and it also improves the prediction function. Sliding a window over the image and characterising the nearby representation for each role into an object/foundation yields article identification. The recognition period is rehashed for scaled versions of the knowledge picture to classify artefacts of different sizes.Finally, window-level recognitions are combined using a mean-shift mode-finding algorithm. The preparation of a classifier for the article identification task is done in two stages. Negative (background) tests are generated in the first stage by randomly separating a set number of features from a background picture arrangement. The final classifier is prepared in a subsequent stage to use both the positive samples, the random negative samples, and the background samples from the previous stage.

4. RESULT COMPARISON

AUTHORS	MODEL	RESULT
Javaan Chahl, Munir Oudah, and Ali-Al Naji	3-D model and depth based model	Accuracy = 99.54% using TOF camera and AVT Martin color camera
Ram Pratap Sharma	HMM and EMG sensors	Accuracy = 95.44%
Zhi-hua Cheng, Jung Tao Kim, Jianning Liang, Jing Zhang and Yo-Bu Yuan	SVM and CRF	Accuracy = 96.69% from 1300ti Accuracy of FEMD = 0.2022
Sara E Garza-Villiare al, Satu E Schaeffer and David J Rios-Soria	SP09, HNW+ 09 and CMO06	Accuracy of left hand = 93.33% Accuracy of right hand = 93.00%
M.I.N.P Munasinghe	Feed-forward neural network	Accuracy in good lighting = 85%, Accuracy in bad lighting = 71.3%

5. CONCLUSIONS

From the overall paper, we approved the performance of ResNet on a freely accessible hand motion dataset. The proposed method performs in a way that is comparable to other commonly used networks, this one is superior. We looked into the number of frames that should be contributing to ResNet. We looked at the number of attention blocks available and where they should be placed in the network. The stacked various soft attention blocks assist the network in perceiving the hand signals with greater precision, according to our findings.

6. REFERENCES

[1]. Nguyen, T.N., Huynh, H.H. and Meunier, J., 2013. " Static hand gesture recognition using artificial neural networks. Journal of Image and Graphics"

[2]. y, G.R.S. and Jadon, R.S., 2009. "A review of vision based hand gestures recognition. International Journal of Information Technology and Knowledge Management

[3]. Omkar Kahane, Jayshree Katkar, Vivek Jadhav,Pratik Jadhav , " Hand Gesture Recognition Approach: A Survey " , From. – International Journal for Innovative Research in Science & Technology , 2017

[4]. Naina Dhingra, Andreas Kunz , " Res3ATN
- Deep 3D Residual Attention Network for Hand Gesture Recognition in Videos " , From. - ETH Zurich, Innovation Center Virtual Reality Leonhardstrasse 21, 8092 Zurich, Switzerland.

[5]. Peter H.N. de With , Rob G.J. Wijnhoven, " Fast Training of Object Detection using Stochastic Gradient Descent " , From. - Univ. of Technol. Eindhoven / CycloMedia Eindhoven, The Netherlands

[6]. Michal Roth and William T. Freeman. " Orientation Histograms for Hand Gesture Recognition " From: IEEE Intl. Wkshp. on Automatic Face and Gesture Recognition, Zurich, June, 1995.

[7]. Naveen Aggarwal, Pragati Garg and Sanjeev Sofat . " Vision Based Hand Gesture Recognition " From : World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:3, No:1, 2009

[8]. S. Park, S. Yu, and J. Kim. "3D hand tracking using Kalman filter in depth space" .

[9]. EURASIP J. Adv. Signal Process. 2012, 36 (2012).

[10]. Rautaray, Siddharth Swarup, and Anupam Agrawal. "A novel human computer interface based on hand gesture recognition using computer vision techniques." In pp. 292-296 in the Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia.2010.)

[11]. Yi, Beifang, Frederick C. Harris Jr, Ling Wang, and Yusong Yan. "Real-time natural hand gestures." Computing in Science & Engineering 7, no. 3 (2005)

[12]. Rautaray, Siddharth Swarup, and Anupam Agrawal. "A vision based hand gesture interface for controlling VLC media player." International Journal of Computer Applications 10, no. 7 (2010)

[13]. Chakraborty, B.K., Sarma, D., Bhuyan, M.K. and MacDorman, K.F., 2017. Review of constraints on vision-based gesture recognition for human–computer interaction. IET Computer Vision, 12(1), pp.3-15.