Real- Time Hand Gesture Recognition Using Deep Learning

Malavika Suresh, Avigyan Sinha, Aneesh R P
Regional Centre, Institute of Human Resources Development, Thiruvananthapuram

Abstract — With the impetuous advancement of informatics, human knowledge is unable to bridge the boundaries and human computer interaction is paving the way for new eras. Here, a real-time human gesture recognition using an automated technology called Computer Vision is demonstrated. This is a type of noncognitive computer user interface, having the endowment to perceive gestures and execute commands based on that. The design is implemented on a Linux system but can be implemented by installing modules for python on a windows system also. OpenCV and KERAS are the platforms used for the identification. Gesture displayed in the screen is recognized by the vision-based algorithms. Using background removal technique, an assortment of skin color masks was trained by Lenet architecture in KERAS for the recognition. The users have tested and produced over 5000 masks with KERAS to generate 96% more accurate results.

Keywords: Gesture Recognition, Computer Vision, Open CV, Lenet, KERAS.

I. INTRODUCTION

Gesture is a sign of physical behavior or an expression of feelings. This involves the movement of the body and the hand. It fell into two categories: [1] [2] static gesture and [3] [4] fluid gesture; for the former, body posture or hand gesture denotes a symbol. For the latter, certain signals are transmitted by body or hand movement. Gesture can be used as computer-human communication tool [5] [6].

Gesture recognition is a part of Human Computer Interaction that shows an academic disquisition and is a key for popularizing the idea of a human-to-man interaction, open dialogue, which must suggest the correlation between the computer and the user [7]. The scientific area of gesture analysis is equipped to identify such gestures as hand, arm, head and even structural motions that typically involve a certain stance and/or motion. The individual can dispatch further specifics in a smaller time span by using hand gestures. Several strategies were used to apply the concepts of computer vision in real-time interpretation of gestures outputs [8].

Here, the Computer vision project focuses on the gesture identification using python language in open CV framework. Language is an enormous communication element. An impaired person cannot communicate with languages. Gesture is an important and relevant method for communication for the impaired ones. So here is a computer based technique by which the ordinary person can grasp what the differently able one tries to convey. There are several similar object recognition systems and algorithms for monitoring. This enables the recognition of gestures to solve the restrictions and limitations prevalent in the previous approaches

II. RELATED WORKS

Numerous robust techniques have been developed and claimed to have accurate performance to tackle gesture recognition. [9] Shows the review of hand gesture recognition system for controlling a robotic arm. Algorithm used for the reconnaissance is neural network, Adaptive Boosting and Support Vector Machine. Hand gestures [10] are implemented by using convex hull for better fingertip detection. The accuracy result for the corresponding paper is more than that of the other existing systems. The objective of [11] is to highlight widely effective methods of capturing gestures which have been fundamental in the recent past.

Manisha U. Kakde, Mahender G. Nakrani, Amit M Rawate, "A Review Paper on Sign Language Recognition for Deaf and Dumb People using Image Processing,", IJERT, ISSN:2278-0181, Vol.5 Issue 03 March 2016: This paper lists current most popular methods of sign acquisition. [12] Paper also shows fingertip recognition using convex hull algorithm by YCbCr color space transformation to detect skin color for contour identification. 10 tested users produce 330 cases of different hand gestures for accurately developed results. [13] Shows a Linux based hand gesture recognition method in python language. Algorithm used is not based on the background

subtraction method for detecting skin color. It recognizes the number of finger tips and as well as the task carried out per requirement.

Kim[14] et al. allowed a user to radio-control a toy car via various hand gestures, recognized from an EMG signal. It listed four movements as appropriate for this mission. Sensors were mounted under the wrist on the lower arm. A combination of naive Bayes and nearest-neighbor classifiers was used for the classification of the gesture. With 30 subjects the method was tested to find the optimal combination of classifiers. User-independent recognition rates ranged between 87% and 98% for this small set of four gestures, showing convincingly that gesture recognition is possible based on such physiological knowledge.

The authors in [15] [16] identify the hand area from input images and then map and evaluate the traveling direction to understand sign language in America. In [17], Shimada et al. using hand gesture recognition to deliver a TV control system. Keskin et al. [18] segment the hand into 21 different regions and train an SVM classifier to model the mutual distribution of these regions for different hand movements in order to identify the actions. Zeng et al. [19] improve the medical service by understanding of hand gestures.

Eirini Mathe, Alexandros Mitsou, Evaggelos Spyrou and Phivos Mylonas, "Hand Gesture Recognition using a Convolution Neural Network" shows hand gesture recognition that uses a Convolution Neural Network (CNN), which is trained on Discrete Fourier Transform (DFT) images that result from a sensor reading. Here they evaluate a dataset of hand gestures involving either one or both hands simultaneously and compare the proposed approach to another that uses hand-crafted features.

III. PROPOSED METHOD

In the following technique, we show how to integrate real-time hand gesture recognition into an advanced technology like Computer Vision. This paper is developed to identify 10 set of gestures shown in the fig below. This tactic was implemented with the Object-oriented python programming language, the Open Source Computer Vision Library, developed by Intel. NumPy libraries were used for the machine learning tasks and for ensuring consistency in large matrix processing tasks. [20] The input is a live webcam for the gesture to recognize and predict.

Studies were conducted under specific and flourishing lighting conditions. Each frame is pre-processed under many conditions before developing mask images from the captured frames.

The methodology proposed can primarily be divided into different parts: image acquisition segmentation, process of classification. The block diagram below shows all the steps performed in the proposed system.

Fig 1: Block Diagram of the Proposed System

IV. METHODOLOGY

A) Image Acquisition

The first phase was to grab the frame from the camera and to establish the region of interest in the image. The challenge is to utilize an adequate computer vision method that can detect the slightest discrepancy between similar signals in order to prevent inaccurate translation in real time. There are over 150 conversion methods in OpenCV for color-space. But only two of them will be examined here, BGR to Gray and BGR to HSV. In this paper we convert the colored RGB format into HSV, as it is more convenient to extract the gesture rather than others [21].

B) Segmentation

Segmentation is also a vital part of the proposed methodology. The process shows a binary object describing the segmentation in the simplistic application. Black pixels are the background pixels and white pixels are the

foreground pixels. In simple implementations, a single parameter called the intensity threshold determines the segmentation. Numerous imperfections can occur in binary images.

[22] In particular, noise and texture distort the binary regions that result from simple thresholds. Morphological image processing explores the purposes of extracting the shape and function of these imperfections [23]. Such methods can be applied to objects in grayscale.

The main 7 processes performed under this operation are erosion, dilation, opening by reconstruction, closing by reconstruction, complement, and regional maxima and superimpose. Erosion and dilation processes are used here to detect the exact gesture from the camera. In this project, over 5000 mask images of the like, unlike, palm, super and fingers from one to five are extracted.

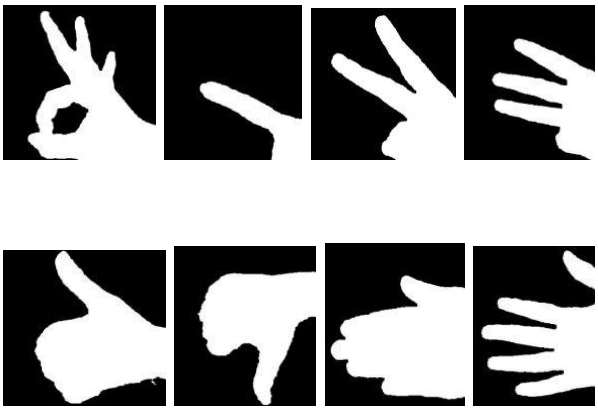

Fig 2: Mask images

C) Classification

The extracted mask images are trained by a convolution neural network Lenet in KERAS, which is a high level API written in python and capable of running on top of tensor flow. A convolution neural network (CNN, or ConvNet) could be a category of deep neural networks in deep learning, most typically used in the visual representation process analysis. CNNs use a multilayer perception variation designed to require the smallest amount of preprocessing [24].

Lenet is a CNN architecture used for the training of the gestures. The architecture of LeNet-5 includes 2 sets of cozy and average pooling layers, then a flattening cozy layer, two fully linked layers and finally a classifier with the softmax framework. In some cases, LeNet-5's main drawback is overflowing and no built-in mechanism to avoid this [25]. By adding dropout layers, the benchmark architecture has been improved [26].

After optimization techniques, the neural network is constructed using the Tensor Flow-based Keras system with the following hyper parameter values:For an individual with limited knowledge of deep learning this can be daunting. KERAS offers an easy and flexible Network Training API that hides most of the complicated information underneath the hood. After training dataset, the output predicts the gesture and shows the accuracy of the image .

IV.RESULTS & DISCUSSIONS

The recognized hand gesture system was tested under various conditions with hand pictures. The methodology details the performance of the entire system using different methods. The limits of the existing systems are overcome by examples of accurate detection and classification [27][28]. Such an insight into the system limitations indicates the direction and focus for future work.

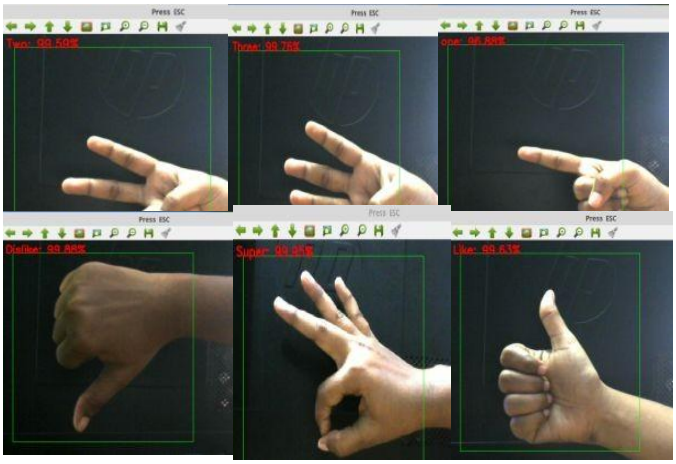

Fig 3: Classification of Gestures

This is an algorithm for classification in real time; users must first train system and then try to recognize the gesture. The webcam acquires hand gesture images and they undergo HSV conversion, filtering of noises by morphological techniques, then gesture extraction of masked frames. Masks are trained by using API KERAS. The training accuracy was obtained above 96% accuracy.

V. CONCLUSION AND FUTURE SCOPE

This experimental hypothesis was intended to identify the movements the user provided to the device with neural network training and classification in real time. More than 5,000 mask pictures in different background have been developed in the hand gesture recognizing method. Grabbed and adapted movements from different people.

The experimental results have shown that the method that can be used for monitoring has a strong identification standard and a low-cost computer technique to recognize the hand of individuals[29].

The vision-based real-time system with Python programming language and OpenCV libraries and Linux framework was implemented effectively. The accuracy of the designed system is 96% and has proven to be perfect for steering in a clear environment. With the implemented system serving as an extendible foundation for the future research, extensions to the current system have been proposed

REFERENCES

[1] A. D. Bagdanov, A. Del Bimbo, L. Seidenari, and L. Usai, "Real-time hand status recognition from RGB-D imagery," in Proceedings of the 21stInternational Conference on Pattern Recognition (ICPR '12), pp. 2456–2459, November 2012.

[2] .M. Elmezain, A. Al-Hamadi, and B. Michaelis, "A robust method for hand gesture segmentation and recognition using forward spotting scheme conditional random fields," in Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10), pp. 3850–3853, August 2010.

[3]. S. Park, S. Yu, J. Kim, S. Kim, and S. Lee, "3D hand tracking using Kalman filter in depth space," Eurasip Journal on Advances in Signal Processing, vol. 2012, no. 1, article 36, 2012.

[4] Y. Wang, C. Yang, X. Wu, S. Xu, and H. Li, "Kinect based dynamic hand gesture recognition algorithm research," in Proceedings of the 4th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC '12), pp. 274–279, August 2012.

[5]. M. Panwar, "Hand gesture recognition based on shape parameters," in Proceedings of the International Conference on Computing, Communication and Applications (ICCCA '12), pp. 1–6, February 2012.

[6]. Z. Y. Meng, J.-S. Pan, K.-K. Tseng, and W. Zheng, "Dominant points based hand finger counting for recognition under skin color extraction in hand gesture control system," in Proceedings of the 6th International Conference on Genetic and Evolutionary Computing (ICGEC '12), pp. 364–367, August 2012.

[7] Swapnil Athavale, Mona Deshmukh, International Journal of Engineering Research and General Science Volume 2, Issue 2, Feb-Mar 2014

[8] Murthy, G.R.S. & Jadon, A review of vision based hand gesture recognition. International Journal of Information Technology and Knowledge Management

[9] Ruchi Manish Gurav P.G." Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV" 2015 International Conference on Industrial Instrumentation and Control (ICIC) College of Engineering Pune, India. May 28-30, 2015

[10] Anupam Agrawal, Rohit Raj and Shubha Porwal "Vision-based Multimodal Human-Computer Interaction using Hand and Head Gestures" IEEE Conference on Information and Communication Technologies ICT 2013.

[11] Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language: A review,", IJSETR, Vol.4, Issue 3, March 2015.

[12] Hsiang-Yueh Lai, Han-Jheng. Lai," Real- Time Dynamic Hand Gesture Recognition" published on 2014 at International Symposium on Computer, Consumer and Control.

[13] Mr. Deepak K. Ray, MayankSoni, PrabhavJohri, Abhishek Gupta, "Hand Gesture Recognition using Python" presented in "International Journal on Future Revolution in Computer Science & Communication Engineering Volume: 4 Issue: 6"

[14] Kim, Jonghwa & Mastnik, Stephan & Andre, Elisabeth. (2008). EMG-based hand gesture recognition for realtime biosignal interfacing. 30-39. 10.1145/1378773.1378778

[15]. R. Yang, S. Sarkar, and B. Loeding, "Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 462–477, 2010.

[16]. Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American sign language recognition with the kinect," in Proceedings of the 13th ACM International Conference on Multimodal Interfaces (ICMI '11), pp. 279–286, November 2011.

[17]. A. Shimada, T. Yamashita, and R.-I. Taniguchi, "Hand gesture based TV control system—towards both user—& machine-friendly gesture applications," in Proceedings of the 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV '13), pp. 121–126, February 2013.

[18] C. Keskin, F. Kiraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV '11), pp. 1228–1234, November 2011.

[19]. J. Zeng, Y. Sun, and F. Wang, "A natural hand gesture system for intelligent human-computer interaction and medical assistance," in Proceedings of the 3rd Global Congress on Intelligent Systems (GCIS '12), pp. 382–385, November 2012.

[20] P.Suganya, R.Sathya, K.Vijayalakshmi, International Journal of Pure and Applied Mathematics, 2018.

[21] Noreen, Uzma & Jamil, Mutiullah & Ahmad, Nazir. (2016). Hand Detection Using HSV Model. Advances in Computer Science and Engineering.

[22] M. Shin, D. Goldgof, and K. Bowyer, An Objective Comparison Methodology of Edge Detection Algorithms using a Structure from Motion Task, in Empirical Evaluation Techniques in Computer Vision, K. Bowyer and P. Philips, IEEE Computer Society Press, Los Alamitos, CA

[23] Srisha, Ravi & Khan, Am. (2013). Morphological Operations for Image Processing : Understanding and its Applications.

[24] Zainab Shukur Mahmood Mahmood, Sait Ali Uymaz,Anatolian Journal of Computer Science,2019.

[25] Ahmed El-Sawy, Mohamed Loey: CNN for Handwritten Arabic Digits Recognition Based on LeNet-5. In book: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics (2016)

[26] Geo_rey E. Hinton, Nitish Srivastava, Krizhevsky A., Sutskever, I., Ruslan R.Salakhutdinov: Improving neural networks by preventing co-adaptation of feature detectors. Department of Computer Science, University ronto, (2012)

[27] Z hong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu, "Object Detection with Deep Learning: A Review",IEEE Publications,2019.

[28] Md Zahangir Alom , Tarek M. Taha ,Chris Yakopcic , Stefan Westberg, Paheding Sidike,Mst Shamima Nasrin , Mahmudul Hasan, Brian C. Van Essen , Abdul A. S. Awwaland Vijayan K. Asari," A State-of-the-Art Survey on Deep Learning Theory and Architectures" Published: 5 March 2019

[29] Yeo, Hui-Shyong & Lee, Byung-Gook & Lim, Hyotaek. (2013). Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. Multimedia Tools and Applicatio