IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

Technology Interaction using Hand Sign using Deep Learning

Dr. Kalaivazhi. V[1], Ajith. M[2], Amal. R[3], Karthickraj. K[4]
Assistant Professor, Anjalai Ammal Mahalingam Engineering College, Kovilvenni[1]
Students, Anjalai Ammal Mahalingam Engineering College, Kovilvenni[2,3,4]

Abstract: *Hand Signs can be used as a human computer interaction tool by computer vision and deep learning. We have designed and developing a system that can identify hand signs and perform appropriate actions in the computer system which are predefined. We will capture the video from web cam and then read the video from image by image. Then the images will be feed into the single shot multiscale boundary detector convolutional neural network to match the pattern it learned which can be folded hand with open thumb and pinky finger in different orientation and fore finger up along with thump or middle finger.*

Keywords*:* Hand Sign Recognition; Camera Sensors; Deep Neural Networks.

I. INTRODUCTION

One might anticipate that as information technology advances in our culture, computer systems will become more integrated into our daily lives. New forms of human-computer interaction with user-friendly interfaces will be required by these environments. Gesture and hand sign recognition has the potential to be a strong and natural tool for efficient and intuitive human-computer interaction.

Every day new applications and devices are becoming part of our life, but the means of communication with these are at moment limited to input devices such as mouse, keyboards, etc. These input devices have grown to be familiar, but they inherently limit the speed and naturalness of our interaction with the computers.

An appealing replacement for bulky interface devices for human-computer interaction is the use of hand gestures as human computer interaction. Users typically utilise hand gestures to describe how they are feeling and to let others know what they are thinking. The comfort and naturalness needed for HCI can be attained, in particular, through the visual interpretation of hand motions.

Deep learning is a subfield of machine learning that involves training artificial neural networks to learn and make predictions from large amounts of data. Unlike traditional machine learning algorithms, which require engineers to manually extract features from the data, deep learning algorithms are capable of automatically learning and extracting features on their own, making them more powerful and versatile.

Deep learning algorithms are composed of multiple layers of interconnected neurons, or nodes, that are capable of processing and transforming data in a non-linear manner. During the training process, these networks adjust the strength of the connections between the neurons to minimize the difference between the predicted outputs and the actual outputs. This is done using optimizationtechniques such as gradient descent, which iteratively adjusts the parameters of the model to improve its accuracy.

With the computer vision technology, we can create a seamless accurate human computer interaction tool with signs of bare hands. This sign detection is achieved by single shot multiscale boundary detector of TensorFlow object detection model and python.

II. RELATED WORK

The earliest gestures used for computer interactions may be traced back to Ivan Sutherland's doctoral thesis, which featured Sketchpad, an early implementation of stroke-based gestures that used a light pen to modify graphical elements on a tablet display. Since then, this gesture has gained a lot of support from the human-computer interface community.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

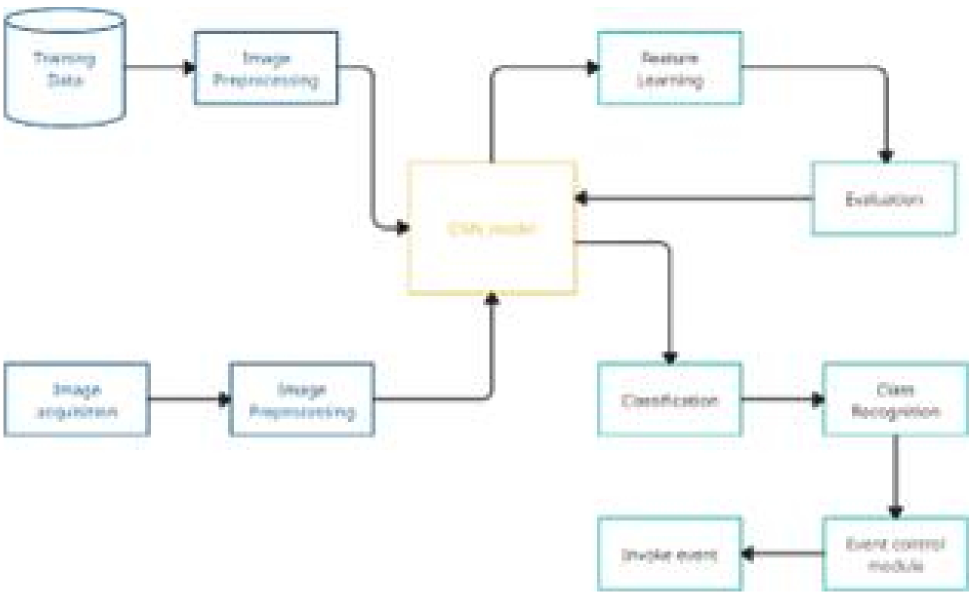
Sign language is the primary mode of communication between hearing and vocally impaired population. The government of India has enacted the Rights of Persons with Disabilities Act 2016 (RPwD Act 2016). This act recognizes Indian Sign Language (ISL) as an important communication medium for communicating with hearing impaired people. This also insists the need for sign language interpreters in all Government organizations and public sector undertakings in order to abide RPwD Act 2016. This can avoid their isolation from the rest of society to a great extent.

In this work, we propose a signer independent deep learning-based methodology for building an Indian Sign Language (ISL) static alphabet recognition system. Here, we review various existing methods in sign language recognition and implement a Convolutional Neural Network (CNN) architecture for ISL static alphabet recognition from the binary silhouette of signer hand region. We also discuss in detail, the dataset used along with the training phase and testing phase of CNN. The proposed method was successfully implemented with an accuracy of 98.64% which is better than most of the currently existing methods.

Hasanuzzaman et al. presented a real-time hand gesture recognition system using skin color segmentation and multiple-feature based template-matching techniques. In their method, the three largest skin-like regions are segmented from the input images by skin color segmentation technique from YIQ color space and they are compared for feature-based template matching using a combination of two features: correlation coefficient and minimum (Manhattan distance) distance qualifier. In their experiment, they have recognized ten gestures out of which two are dynamic facial gestures. These Gesture commands are being sent to their pet robots AIBO through Software Platform for Agent and Knowledge Management (SPAK) and their actions are being accomplished according to users pre-defined action for that gesture.

Nielsen et al. proposed a real time vision system which uses a fast segmentation process to obtain the moving hand from the whole image and a recognition process that identifies the hand posture from the temporal sequence of segmented hands. The system 's visual memory stores all the recognizable postures, their distance transform, their edge map and morphologic information. They have used Hausdorff distance approach for robust shape comparison. Their system recognitions 26 hand postures and achieved a 90% recognition average rate.

III. SYSTEM ARCHITECTURE



A. Data Collection:

The data collection is the first process of training and testing of deep learning model. It is carried out by collecting images from sensors like web cam of the laptop or from hand held devices.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 6, May 2023

B. Data Pre-processing:

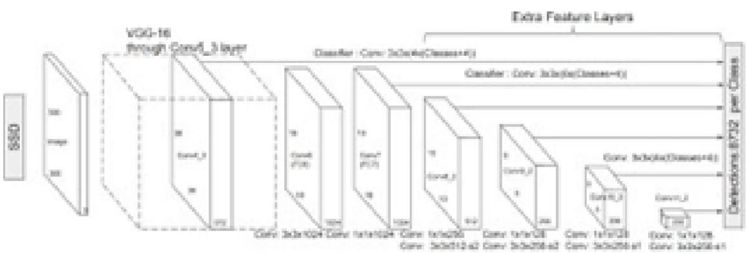
The data pre-processing involves modifying the data into a form that model accepts. That is the captured image is converted into the shape of 225x225 which will be again converted to xml file that is the type of file the single shot multiscale detector model accepts.

C. Model Development:

The model used by as is single shot multiscale boundary detector (SSD) convolutional neural network which uses the architecture of VGG16 which is powerful model that is faster than YOLO object detection. The model is tuned before training and testing. The tuning involves unfreezing certain layers of the model and adding the required classification layers on the top with activation of softmax.

$$s\left(x_i\right) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

The single shot multiscale detector uses VGG16 architecture which involves 13 convolutional layers, 5 pooling layers and 3 dense layers as default. Along with these layers, one softmax multiclass classification layer and two average pooling layers will be added at the top of the VGG16 neural net. The single shot detector architecture is,



Then the model will be trained on the training set of 70% and then evaluated with the testing set of 30% from the collected data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

D. Prediction Result:

The result will be measured in metric called Accuracy which shows the model's performance in 0 – 100 %. The developed model shows the accuracy of 89 % for most case and it will return the result to the event control module to invoke the appropriate event in the system. Such as if the event controller invokes the hand mouse module the result will be controlling of cursor it involves hand detection and coordinate matching by which we can move the mouse cursor to where ever we want in the monitor and if it invokes the volume controller module it will adjust the volume by 25%, 50%, 70% and 100% based on the sign detected.

IV. CONCLUSION

We developed an advance sign detection model with single shot multiscale boundary detector and object detection models. In future, this can be further improved to multiple sign detection in less computational power. And we can add many more events in the modules with different types of gesture signs.

REFERENCES

[1] S. Mitra and T. Acharya. Gesture recognition: A survey. IEEE Systems, Man, and Cybernetics, 37:311–324, 2007.

[2] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human- computer interaction: A review. PAMI, 19:677–695, 1997.

[3] P. Trindade, J. Lobo, and J. Barreto. Hand gesture recognition using color and depth images enhanced with hand angular pose data. In IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, pages 71–76, 2012.

[4] J. J. LaViola Jr. An introduction to 3D gestural interfaces. In SIGGRAPH Course, 2014.

[5] T. Starner, A. Pentland, and J. Weaver. Real-time American sign language recognition using desk and wearable computer-based video. PAMI, 20(12):1371–1375, 1998.

[6] S. B. Wang, A. Quattoni, L. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. In CVPR, pages 1521–1527, 2006.

[7] N. Dardas and N. D. Georganas. Real-time hand gesture detection and recognition using bag- of-features and support vector machine techniques. IEEE Transactions on Instrumentation and Measurement, 60(11):3592–3607, 2011.

[8] M. Zobl, R. Nieschulz, M. Geiger, M. Lang, and G. Rigoll. Gesture components for natural interaction with in-car devices. In Gesture-Based Communication in Human Computer Interaction, pages 448–459. Springer, 2004.

[9] F. Althoff, R. Lindl, and L. Walchshausl. Robust multimodal hand-and head gesture recognition for controlling automotive infotainment systems. In VDI-Tagung: Der Fahrerim 21. Jahrhundert, 2005.

[10] F. Parada-Loira, E. Gonzalez-Agulla, and J. Alba-Castro.Hand gestures to control infotainment equipment in cars. In IEEE Intelligent Vehicles Symposium, pages 1–6, 2014.

[11] D. C. Cires¸an, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In International Joint Conference on Artificial Intelligence, pages 1237–1242, 2011.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In NIPS, pages 1097–1105. 2012.

[13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradientbased learning applied to document recognition. In Proceedings of the IEEE, pages 2278–2324, 1998.

[14] P. Y. Simard, D. Steinkraus, and J. C. Platt. J.c.: Best practices for convolutional neural networks applied to visual document analysis. In Int. Conference on Document Analysis and Recognition, pages 958– 963, 2003.

[15] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In CVPR, pages 3642–3649, 2012.

[16] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In CVPR, pages 1725–1732, 2014.

[17] L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen. Sign language recognition using convolutional neural networks. In ECCVW, 2014.

[18] K. Simonyan and A. Zisserman. Two-stream convolutional networks for action recognition in videos. In NIPS, pages 568–576, 2014.
[19] P. Molchanov, S. Gupta, K. Kim, and K. Pulli. Multi-sensor System for Driver's Hand-gesture Recognition. In AFGR, 2015.
[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International Conference on Machine Learning, 2015, pp. 448–456.
[21] Raymond Ahn, Justin Zhan, Using proxies for node immunization identification on large graphs, IEEE Access, Vol. 5, pp. 13046-13053, 2017.
[22] Gary Blosser, Justin Zhan, Privacy preserving collaborative social network, International Conference on Information Security and Assurance, pp. 543-548, 2008.
[23] C Chiu, J Zhan, F Zhan, Uncovering suspicious activity from partially paired and incomplete multimodal data, Vol. 5, pp. 13689-13698, IEEE Access, 2017.
[24] Brittany Cozzens, Richard Huang, Maxwell Jay, Kyle Khembunjong, SahanPaliskara, Felix Zhan, Mark Zhang, Shahab Tayeb, Signature Verification Using a Convolutional Neural Network, University of Nevada Las Vegas AEOP/STEM/REAP/RET Programs Technical Report, 2018.