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 14, May 2023

Real Time Gesture and Object Detection

Pallavi Arote, Akanksha Gaikwad, Rohini Gorhe, Manisha Shete
Matoshri College of Engineering and Research Centre (MCOERC), Nashik, Maharashtra, India

Abstract: *The project goal is to produce a working system for detecting the objects from given video source. In critical situation this detection will help us in finding a way. In this we will be using different cameras for detection of object. Calculates different parameters from the objects from the objects in the field of computer vision. The problems of detect single object and detect multiple object are not same..For detecting multiple objects, lots of problem can arise due to adrupt object motion, multiple object interaction , drifting of object etc. The main goal of this project is to help deaf and dumb people to communicate with normal people and also it helps normal people to understand their sign language. Through this application deaf and dumb people can easily communicate with normal people.*

Keywords*:* Sign Language Recognition

I. INTRODUCTION

Sign language is largely used by the disabled, and there few others who understand it, such as reatives activity, and teachers. Natural gesture and formal cues are the two types of signlanguage.The natural cue is a manual expressiom agreed upon by the user , recognized to be limited in a particular group (esoteric), and a substitute for words used by a deaf person (as opposed to body language). A formal gesture is a cue that is established deliberately and has the same language structure as the community's spoken language More than 360 million of world population suffers from hearing and speech impairments. Sign language detection is a project implementation for designing a modelin which web camera is used for capturing images of hand gestures which is done by open cv. After capturing images, labelling of images are required and then pre trained model SSD Mobile net v2 is used for sign recognition. Thus, an effective path of communication can be developed between deaf and normal audience. Three steps must be completed in real time to solve our problem:

1. Obtaining footage of the user signing is step one (input).
2. Classifying each frame in the video to a sign
3. Reconstructing and displaying the most likely Sign from classification scores (output).

II. RELATED WORK

With the continuous development in Information technology the ways of interaction between computers and Humans have also evolved. There has been a lot of work done in this field to help deaf and able-bodied people communicate more effectively. Because sign language is a collection of gestures and postures, any effort to recognise sign language falls under the purview of human computer interaction.

Sign Language Detection is categorised in two parts. The first category is the Data Glove approach, in which the user wears a glove with electromechanical devices attached to digitise hand and finger motion into processable data.

The disadvantage of this method is that you must always wear extra gear and the results are less accurate. In contrast, the second category, computer-vision-based approaches, require only a camera, allowing for natural interaction between humans and computers without the use of any additional devices. Apart from various developments in ASL field, Indian people started putting work in ISL. Like Image key pointdetection using SIFT, and then comparing the key point of a new image to the key points of standard images per alphabet in a database to classify the new image with the label of the closest match. Similarly various work has been put into recognising the edges efficiently one of the idea was to use a combination of the colour data with bilateral filtering in the depth images to rectify edges. With Advancement in Deep Learning and neural networks people also implementing them in improving the detection system.

In reference , the ASL is recognised using a variety of feature extraction and machine learning techniques, including the Histogram technique, the Hough transform, OTSU's segmentation algorithm, and a neural network.

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 14, May 2023

The head subpart will be further categorized into pose and movements as well as facial expressions. Postures and gestures will be extracted from the movement of the hands. All of the data will then be matched against the WLASL Dataset which would then be used for classification purposes. The classification will result in generation Of words. Words generated from sign language will not adhere to grammatical rules of English. Hence semantically correct sentences will be generated by the sentence generation module. For this Google's T5 model [19] will be put into use. Finally, this output will be sent through an audio generator to generate speech from the same. This provides support for illiterate people, who are not entitled to understanding written text.

Image processing is concerned with computer processing the images which include collecting, processing, analysing and understanding the results obtained. Computer vision necessitates a combination of low-level image processing to improve image quality (e.g., removing noise and increasing contrast) and higher-level pattern recognition and image understanding to recognise features in the image

III. METHODOLOGY

Our project aims to capture sign language performed by signers on a real-time basis and interpret the language toproduce textual and audio output for the illiterate. For this, acamera-based approach will be made use of, ease of portability and movement that the camera-based method offers over other techniques.

The video of the signer will be first captured by a camera-enabled device.This video will then be processed our application. The video would be divided into a Numberof frames which will convert the video into a raw image sequence. This image sequence will then be processed to initially identify the boundaries. This will separate the different body parts being captured by the camera into two major subparts - head and hands.

All of the data will then be matched against the WLASL Dataset which would then be used for classification purposes. The classification will result in generationOf words. Words generated from sign language will not adhere to grammatical rules of English. Hence semantically correctsentences will be generated by the sentence generation module. For this Google's T5 model [19] will be put intouse. Finally, this output will be sent through an audiogenerator to generate speech from the same. This providessupport for illiterate people, who are not entitled to understanding written text.

IV. RESULTS

The model was trained using the technique of transfer learning and a pre-trained model SSD mobile net v2 was used

Transfer Learning:

Transfer learning is a concept that describes a process in which a model that has been trained on one problem is applied in some way to a second, related problem. Transfer learning is a deep learning technique that includes training a neural network model on an issue that is similar to the one being addressed before applying it to the problem at hand. Using one or more layers from the learnt model, a new model is then trained on the problem of interest.

SSD Mobile net V2:

The Mobile Net SSD model is a single-shot multibox detection (SSD) network that scans the pixels of an image that are inside the bounding box coordinates and class probabilities to conduct object detection. In contrast to standard residual models, the model's architecture is built on the notion of inverted residual structure, in which the residual block's input and output are narrow bottleneck layers. In addition, nonlinearities in intermediate layers are reduced, and lightweight depthwise convolution is applied. The TensorFlow object detection API includes this model.

V. APPLICATION AND FUTURE SCOPE

Application:

- The dataset can easily be extended and customized according to the need of the user and can prove to be an important step towards reducing the gap of communication for dumb and deaf people.
- Using the sign detection model, meetings held at a global level can become easy for the disabled people to understand and the value of their hard work can be given.
- The model can be used by any person with a basic knowledge of tech and thus available for everyone.

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 14, May 2023

- This model can be implemented at elementary school level so that kids at a very young age can get to know about the sign language.

Future scope:

- The implementation of our model for other sign languages such as Indian sign language or American sign language.
- Further training the neural network to efficiently recognise symbols.
- Enhancement of model to recognise expressions.

REFERENCES

[1] D. Metaxas. Sign language and human activity recognition, June 2011. CVPR Workshop on Gesture Recognition.
[2] M. Ranzato. Efcient learning of sparse representations with an energy-based model, 2006. Courant Institute of Mathematical Sciences.
[3] S. Sarkar. Segmentation-robust representations, matching, and modeling for sign language recognition, June 2011. CVPR Workshop on Gesture Recognition, Co-authors: Bar-bara Loeding, Ruiduo Yang, Sunita Nayak, Ayush Parashar.
[4] P. Y. Simard. Best practices for convolutional neural networksapplied to visual document analysis, August 2003. Seventh International Conference on Document Analysis and Recog nition.
[5] X. Teng. A hand gesture recognition system based on local linear embedding, April 2005. Journal of Visual Language September 2018.
6] Chen L, Lin H, Li S (2012) Depth image enhancement for Kinect using region growing and bilateral filter. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012). IEEE, pp 3070–3073
[7] Vaishali.S.Kulkarni et al., "Appearance Based Recognition of American Sign Language Using Gesture Segmentation", International Journal on Computer Science and Engineering (IJCSE), 2010
[8] Cheok, M. J., Omar, Z., &Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10(1), 131-153
[9] Al-Saffar, A. A. M., Tao, H., & Talab, M. A. (2017, October). Review of deep convolution neural network in image classification. In 2017 International Conference on Radar, 37 International Journal for Modern Trends in Science and Technology Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) (pp. 26-31). IEEE.
[10] Kiron Tello O'Shea, An Introduction to Convolutional Neural Networks, (2015 November). Research GATE
[11] https://www.exastax.com/deep-learning/top-five-use-cases-of-tensorflow/
[12] https://en.m.wikipedia.org/wiki/OpenCV
[13] https://github.com/tzutalin/labelImg
[14] [Page 538 https://www.amazon.com/Deep-Learning-Adaptive-Computation-, Deep Learning, 2016.]
[15] https://machinethink.net/blog/mobilenet-v2/ by Matthijs