

# **SIGN LANGUAGE TRANSLATOR USING MACHINE LEARNING**

Vishwas S, Hemanth Gowda M, Vivek Chandra H N, Tannvi

Department of Computer Science & Engineering

Vidyavardhaka College of Engineering

Mysuru, Karnataka

Email: vishwas.s.gurkar@gmail.com

hemanthgowda1996@gmail.com

vivekunemployed@gmail.com

komal.tanvi72@gmail.com

Dr Ravi Kumar V

Head of the Department of Computer Science & Engineering

Vidyavardhaka College of Engineering

Mysuru, Karnataka.

Email: ravikumarv@vvce.ac.in

*Abstract* - Sign language is an incredible advancement that has grown over the years. Unfortunately, there are some drawbacks that have come along with this language. Not everyone knows how to interpret a sign language when having a conversation with a deaf and dumb person. There is always a need to communicate using sign language. One finds it hard to communicate without an interpreter. To solve this, we need a product that is versatile and robust. We need to convert the sign language so that it is understood by common people and will help them to communicate without any barriers. The main purpose of this project is to eliminate the barrier between the deaf and dumb and the rest.

*Key Words*– Pose estimation, decision trees, and sign language

## **1. INTRODUCTION**

Machine learning provides a versatile and robust environment to work on. The machine learning subject also eliminates the need for the coder to write updates whenever a new sign is read, this will be done by the machine itself.

Our system aims to get the deaf and dumb people more involved to communicate and the idea of a camera-based sign language recognition system that would be in use for converting sign language gestures to text (English) and then to regional languages. Our objective is to design a solution that is intuitive and simple which simplifies the communication for the majority of people with deaf and dumb people.

There are many methods to convert the sign language which often use Kinect as the basic system to get the inputs and work on them for conversion. Kinect methods are complicated in so many aspects. Our approach will be simple. We use simpler ways to capture the inputs and process them. We have used common and easily available libraries in our system.

## 2. RELATED WORK

A concept of 2D Pose Estimation using Part Affinity Fields signified us about the Human 2D pose estimation—the problem of localizing anatomical key points [1][2][3]. Pictorial Structures for Object Recognition is another topic which highlighted the representation of an object by a collection of points arranged in a formable configuration [4][5]. Monocular 3D pose estimation and tracking by detection signified about 3D pose estimation and tracking of multiple people in cluttered scenes using a monocular, potentially moving camera. One more concept on 2D human pose estimation, a new benchmark and state of the art analysis highlights about articulated human pose estimation using a new large-scale benchmark dataset [6][7]. Strong appearance and expressive spatial models for human pose estimation topic signifies about the requirement of 17 points in pose estimation [8].

Decision tree algorithm optimization research based on MapReduce topic viewed upon an optimized genetic algorithm which is merged into the implementation of the decision tree algorithm [9] [10]. A Survey on Decision Tree Algorithm for Classification helped us with various algorithms of Decision tree, their characteristic, challenges, advantage and disadvantage [11] [12]. The topics of decision tree induction using machine learning [13] and failure analysis [14].

‘India's first-ever sign language dictionary’ by Indian Sign Language Research and Training Centre (ISLRTC) is a collection of more than 3000 words [15]. Google-trans Documentation Release 2.2.0 will be our reference for translation part.

## 3. PROPOSED APPROACH

The main objective of this project is to recognizing the gestures and displaying the correspondent word. The first phase involves capturing the gesture using a webcam along with pose estimation library [1]. The webcam captures the image and image is processed with pose estimation algorithm in tensor-flow utility. Fig 2 shows how the webcam is reading the image and the skeleton mapped on the image is the result of the pose estimation library. The skeleton obtained provides the values for creating the data set; the data set is a collection of the values of the coordinates of the end points of the skeleton. These values are labelled accordingly and are appended to the machine for predicting [13] when the input is taken. The block diagram in Fig 1 explains how the work is carried out in the system.

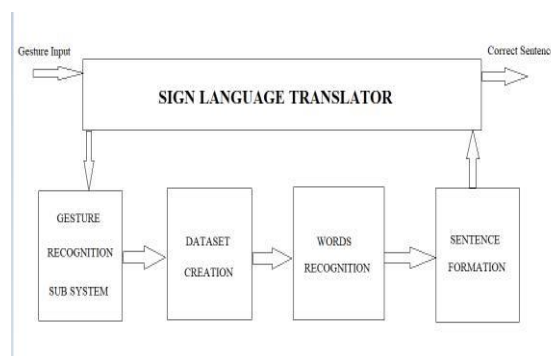


Fig 1: Block diagram of the model

### 3.1 EXTRACTION OF GESTURES

Capturing signs from real world and translating them is the core objective of this work. The real-world signs are read using a webcam which captures both static and moving images of the objects in front of it. The deaf and dumb person who is signing is made to stand in front of the webcam and the image captured from this is processed with the tf-pose-estimation [1] [2] [3] library to map out the skeleton of the person signing. Fig 2 is an example of how the skeleton is mapped on the system.

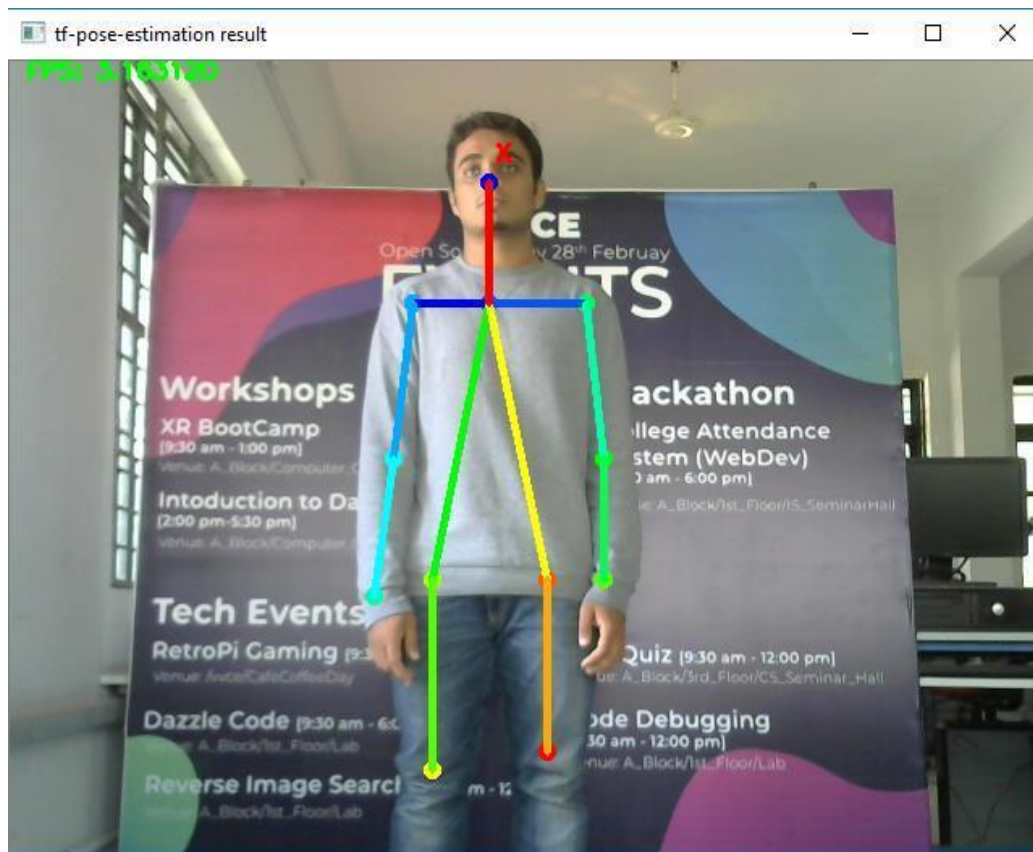


Fig 2: Tf-pose-estimation result

Tf-pose-estimation basically sketches out a stick figure of the body. When the webcam is running the pose estimation algorithm identifies the key points on the subject's body such as elbow joints, wrist, knee joints etc and connects them as one skeleton. The key points namely the end points of the skeleton are labelled with x, y and z co-ordinates for every frame captured. As such 17 [4] [8] key points are identified from the pose-estimation algorithm. The value of these coordinates change for different gestures and the relative distance between the key points is different for different people (as size changes from person to person). These coordinates are the main component to form the data set for training.

### 3.2 CREATING DATA SET

Each gesture captured has its coordinates values stored in a csv (comma separated file) file and the corresponding labels are written in another csv file. Fig 3 shows how a csv file is written and Fig 4 shows the values of csv file in excel.

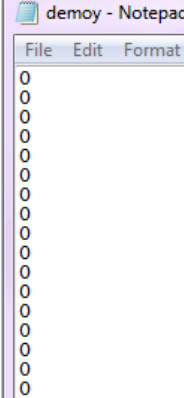
```
6.921656643555956556e+00,-1.229869805307162096e+02,-1.525415161313897272e+02,-1.00932580251
00283615083098994e+02,2.053708466484514759e+02,4.417152439613985280e+02,5.90016237571724445
7.194720296892704425e+00,-1.161306281718584614e+02,-1.417762758387786448e+02,-9.75893828333
680786685216186243e+02,2.050572213370666930e+02,4.556006470602119407e+02,5.9432321460049433
1.380670645997282620e+01,-1.097046183854091908e+02,-1.385871497647455897e+02,-8.81721058800
.641141339429797199e+02,2.085976645909441061e+02,4.496677840573361777e+02,5.856233484013878
7.474693382577323852e+00,-1.147716475977152299e+02,-1.077250039328618669e+02,-8.52940783511
773974716476019466e+02,2.066448082792872754e+02,4.562027337581357074e+02,5.9559166057389199
1.286248678446395388e+01,-1.110455801590472191e+02,-1.398338155749669056e+02,-9.01066445153
675564303527426091e+02,2.054250977939866516e+02,4.544999246719443704e+02,5.9245250460286581
7.194720296892704425e+00,-1.161306281718584614e+02,-1.417762758387786448e+02,-9.75893828333
680786685216186243e+02,2.050572213370666930e+02,4.556006470602119407e+02,5.9432321460049433
6.243136453014972886e+00,-1.173224238247998841e+02,-1.431371706036606213e+02,-1.00059410837
714667995285735742e+02,2.052264153199473924e+02,4.605113491448474292e+02,5.9531565976064564
1.292054396335170097e+01,-1.108200451850583050e+02,-1.398815790285921992e+02,-9.05275327897
```

Fig 3: CSV files for the coordinates

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	6.92E+00	-1.23E+02	-1.53E+02	-1.01E+02	1.28E+02	1.11E+02	1.13E+02	1.39E+01	1.40E+01	-1.49E+01	-1.57E+01	2.06E+02	2.82E+02	2.36E+02	-1.75E+02	-2.72E+02
2	7.19E+00	-1.16E+02	-1.42E+02	-9.76E+01	1.39E+02	1.27E+02	1.13E+02	1.03E+01	-3.53E+00	-1.06E+01	-6.62E+00	1.91E+02	2.65E+02	2.55E+02	-1.70E+02	-2.69E+02
3	1.38E+01	-1.10E+02	-1.39E+02	-8.82E+01	1.42E+02	1.28E+02	1.26E+02	1.65E+01	-5.43E+01	-1.49E+01	-4.90E+00	1.92E+02	2.64E+02	2.17E+02	-1.65E+02	-2.62E+02
4	7.47E+00	-1.15E+02	-1.08E+02	-8.53E+01	1.39E+02	1.30E+02	1.15E+02	8.28E+00	-3.44E+00	-5.47E+00	-5.49E+00	1.90E+02	2.60E+02	2.17E+02	-1.69E+02	-2.68E+02
5	1.28E+01	-1.11E+02	-1.40E+02	-9.01E+01	1.43E+02	1.28E+02	1.21E+02	1.69E+01	-2.04E+01	-1.43E+01	-4.58E+00	1.94E+02	2.66E+02	2.19E+02	-1.65E+02	-2.63E+02
6	7.19E+00	-1.16E+02	-1.42E+02	-9.76E+01	1.39E+02	1.27E+02	1.13E+02	1.03E+01	-3.53E+00	-1.06E+01	-6.62E+00	1.91E+02	2.65E+02	2.55E+02	-1.70E+02	-2.69E+02
7	6.24E+00	-1.17E+02	-1.43E+02	-1.00E+02	1.40E+02	1.28E+02	1.13E+02	9.89E+00	-3.21E+00	-9.29E+00	-6.07E+00	1.93E+02	2.68E+02	2.57E+02	-1.71E+02	-2.72E+02
8	1.29E+01	-1.11E+02	-1.40E+02	-9.05E+01	1.43E+02	1.29E+02	1.26E+02	1.61E+01	-2.03E+01	-1.36E+01	-4.32E+00	1.94E+02	2.67E+02	2.20E+02	-1.66E+02	-2.64E+02
9	8.21E+00	-1.15E+02	-1.41E+02	-9.61E+01	1.39E+02	1.27E+02	1.10E+02	1.00E+01	-3.70E+00	-1.11E+01	-6.88E+00	1.90E+02	2.63E+02	2.54E+02	-1.69E+02	-2.67E+02
10	8.21E+00	-1.15E+02	-1.41E+02	-9.61E+01	1.39E+02	1.27E+02	1.10E+02	1.00E+01	-3.70E+00	-1.11E+01	-6.88E+00	1.90E+02	2.63E+02	2.54E+02	-1.69E+02	-2.67E+02
11	8.41E+01	-1.22E+02	-1.09E+02	-9.71E+01	1.36E+02	1.32E+02	1.00E+02	-2.43E+00	-7.11E+00	4.33E+00	-6.33E+00	1.89E+02	2.61E+02	2.59E+02	-1.77E+02	-2.80E+02

Fig 4: Excel view of the csv file

Each frame as 17 key points and each of these points has 3 coordinates; therefore, there exists 17x 3 values [5] [7] for each frame. This entire set of 51 values is labelled together as one gesture. We are assigning numbers for these gestures in the corresponding csv file as it is easy to handle integers while training the machine. Fig 5 shows how the labelling is done in csv files.



	A
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0

Fig 5: csv and txt file representation of the labels

Here '0' is the uses to label certain gestures for example in Fig 2 a boy standing 'Idle' is recorded Fig 3 and Fig 4 are the representation of the coordinates of his skeleton, and '0' is the label for this posture. Later '0' is substituted for "Idle" while displaying the result. Training the machine requires several sets of values, therefore a single person has to record many frames for a single gesture and the same gesture has to be signed by different people. Different people are required to sign for the same posture as the size of the skeleton varies from person to person. Fig 6 shows how different the skeleton size is for different people. Several frames have to be recorded for a single gesture by a single person and several people have to record the same gesture to provide a better data set for training [13] [14].

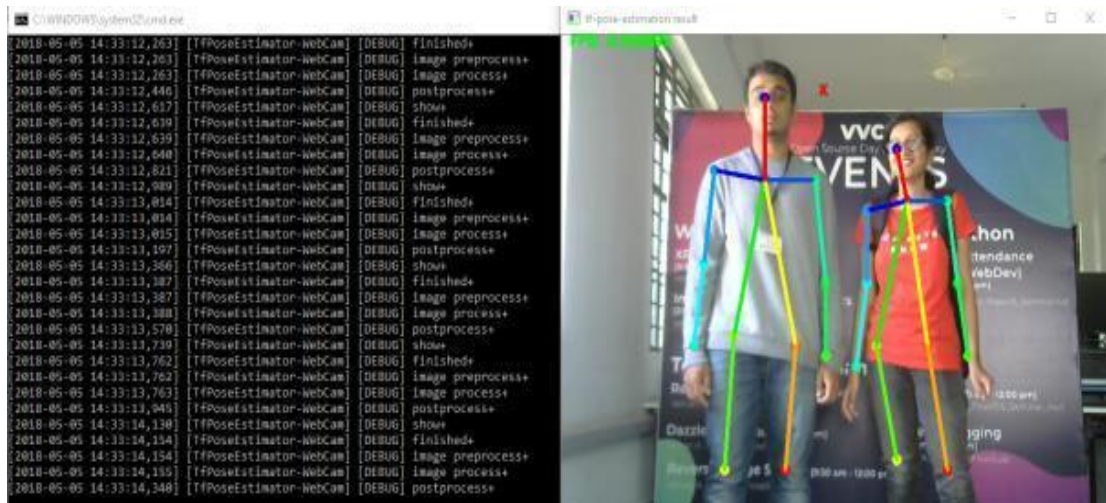


Fig 6: Different skeleton size for two different people

### 3.3 TRAINING

The data set created is taken up in the training platform and Decision tree algorithm [9] [10] is used to train the machine. The set of values stored for a specific gesture is referred by the machine in its training and makes it possible to predict the gesture when the input is taken. Working of the decision tree is shown in Fig 7.



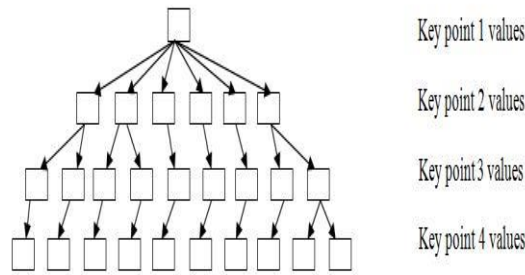


Fig 7: Example of how decision tree is using the values of the key points

The input values will be run through the tree and the final answer will be displayed along with its value and the corresponding label. The corresponding values for labels are then substituted with words and are displayed in the result. Every new gesture has several frames recorded for it and trained using the decision tree algorithm [13]. More the number of frames recorded better the efficiency of the system in predicting the gesture [11] [12].

#### 4. RESULTS

The system, when provided with the proper gestures, gives out the corresponding words. The system can provide proper results even when there are some slight variations in gestures. There will be different kinds of variations from different kinds of persons performing the gestures. The system recognizes multiple gestures one after the other and gives out the respective words. The overall system performance is shown in the Table 1.

Table 1: System Accuracy

Gestures	Accuracy	False Negative	False Positive
Stomach-ache	80%	20%	30%
Headache	80%	20%	30%
Dancing	70%	30%	20%
Studying	70%	30%	20%

#### 5. CONCLUSION AND FUTURE ASPECTS

The requirement of machine-based sign language translator is very important in the present scenario. Even though we have found initial success in this regard, lot of work needs to be done.

- The main area where this can be used is in public places like ticket issuing counters, hospitals etc.
- This can be even used to teach the sign language to normal people.
- Further this can be used to take words and display the gesture for the same.
- Recognizing fingers will widen the training set for the machine.

## REFERENCES

- [1] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh. 'Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields'. The Robotics Institute, Carnegie Mellon University on 14 April 2017.
- [2] V. Belagiannis and A. Zisserman. Recurrent human pose estimation. In 12th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2017.
- [3] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy. Towards accurate multi-person pose estimation in the wild. arXiv preprint arXiv:1701.01779, 2017. In CVPR, 2017.
- [4] Pedro F. Felzenszwalb, Daniel P. Huttenlocher. 'Pictorial Structures for Object Recognition'. AI Labs MIT, CSE, Cornell University.
- [5] N.J. Ayache and O.D. Faugeras. A new approach for the recognition and positioning of two-dimensional objects. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(1):44–54, January 1986.
- [6] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, Bernt Schiele. 2D human pose estimation: New benchmark and state of the art analysis. In CVPR, IEEE Conference on 25th September 2014.
- [7] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, Bernt Schiele. Strong appearance and expressive spatial models for human pose estimation. In ICCV, IEEE International Conference on 1-8 December, 2013.
- [8] S. Ioffe and D.A. Forsyth. Probabilistic methods for finding people. International Journal of Computer Vision, 43(1):45–68, June 2001.
- [9] Fangfang Yuan, Fusheng Lian, Xingjian Xu. Decision tree algorithm optimization research based on MapReduce. ICSESS, 6th IEEE International Conference on 23-25 September, 2015.
- [10] Chen Jin, Luo De-lin and mu Fen-xiang. An improve ID3 Decision tree algorithm. IEEE 4th International Conference on computer Science & Education.
- [11] Mr. Brijain R Patel, Mr. Kushik K Rana. A Survey on Decision Tree Algorithm for Classification, IJEDR, 2014'.
- [12] Gordan.V.Kass(1980). An exploratory Technique for investigation large quantities of categorical data Applied Statics, vol 29, No .2, pp. 119-127.
- [13] Quinlan J. R. (1986). Induction of decision trees. Machine Learning, Vol.1-1, pp. 81-106.
- [14] M. Chen, A. Zheng, J. Lloyd, M. Jordan and E. Brewer (2004). Failure diagnosis using decision trees. Proc. of the International Conference on Autonomic Computing.
- [15] 'India's first-ever sign language dictionary' by Indian Sign Language Research and Training Centre (ISLRTC).

## **Authors Biography**

We are the students of department of Computer Science and Engineering in Vidyavardhaka College of Engineering, Mysore. We have done this project as part of our final year major project. Our project guide is Dr Ravi Kumar V, Head of the Department of CSE in VVCE, Mysore. We were constantly guided by him throughout the project completion process. Our guide was of immense help to us in cracking whatever problems we faced while completing this project. We were also helped by many faculties of our department in various stages of our project. We four of us are also classmates since the first year and we share a good rapport among ourselves. All these made us to complete this project in much easier way.



VISHWAS S



VIVEK CHANDRA H N



HEMANTH GOWDA M



TANNVI



DR RAVI KUMAR V