International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

A Real-Time Ethiopian Sign Language to Audio Converter

Yigremachew Eshetu
Electrical and Computer Engineering
Wolaita Sodo University, Ethiopia

Endashaw Wolde
Computer Science
Wolaita Sodo University, Ethiopia

***Abstract*— Communication is an important and vital human characteristic in order to experience the human nature of being a social animal. The whole main points in human society like political, social, economic interaction directly depends on communication. In everyday life, there are a variety of communications patterns; with work colleagues, family, neighbors, friends and etc. A different part of different human society uses different types of communication means when we say different part of the society based on the social characteristics, culture lifestyle and etc. A communication means used in the expression of lifestyle, culture, point of view. But when we come to the hearing and speech-disabled part of the society while sharing this moral, culture, and lifestyle with the rest of the society, sadly they cannot experience any of it because of a lack of communication tool between them. This is a communication gap between the deaf and the hearing part of the society.**

There is an estimated amount of 250000 – 1,000,000 deaf sign language users in Ethiopia. This data shows how many people are affected by this communication barrier from the different aspect from their social interaction, getting service from different service centers, education access, political participation and a lot more that consume time to mention. In general, this has been the elephant in the room that is almost ignored by the society or not given enough attention by the concerning sector to bring out a solution. So to tackle this bottleneck or to make a bridge over this communication barrier we have proposed a sign language translator system that can be an intermediate between the deaf and the hearing to solve this communication barrier. This platform uses high-bird based approach and uses machine learning and single-shot detection or convolutional neural network to recognize and translate the Ethiopian Sign Language into speech or text.

Keywords—Ethiopian Sign language; Machine Learning; Neural Network; Deaf; Communication.

I. INTRODUCTION

A different part of different human society uses different types of communication means when we say different part of the society based on the social characteristics, culture lifestyle and etc. A certain language is usually used as a means of communication by human beings. It is essentially a means of communication among the members of society. In the expression of lifestyle, culture, point of view and etc. language plays a vital role in society. It is the tool that conveys traditions and values related to group identity. But when we come to the hearing and speech-disabled part of the society while sharing this moral, culture, and lifestyle with the rest of the society, sadly they cannot experience any of it because of a lack of communication tool between them. there are an estimated amount of about 9.9 – 14.9% of our population has Hearing & Speaking Disabilities and 17.3-22.3% of the population has hearing Problem [1].

The 2013 edition of Ethnologue lists 137 sign languages [2]. But in case of Ethiopian Sign Language, presumably, a national standard sign language used among the deaf. We think the system we have developed enables to bridge over the communication barrier using a real-time Sign Language Recognition (SLR) and Sign Language Translation (SLT).

II. BACKGROUND

The study made in 1994, about 5.9% of Ethiopia's population has Hearing & Speaking Disabilities and 13.3% of the population has hearing Problem [3]. Even if we don't have the current data to know the number or percentage of hearing and speech impaired people in Ethiopia we can certainly say that the percentage goes up by 4 to 9 percent in the last 16 years as the population increases from 62.8 million in 2007 to 104 million in 2018 showing a growth by 65.6%. So using our prediction system we can say that about 9.9 – 14.9% of our population has Hearing & Speaking Disabilities and 12.3- 17.3% of the population has hearing Problem.

There is no efficient and accessible way to communicate, to socialize, to work together, to get service or give service and so much more. This creates a big communication barrier between the hearing and speech-disabled people and the rest of society. This communication-barrier generally creates a separate world for the disabled (hearing and speech impaired) where the two parts try not to face each other in any situation, not in an obvious way but deep down almost nobody wants to be in that situation. From these numerous of sectors where this creation of the Two-world affect both the disabled and the non-disabled some of them are like in a school the disabled have to go to a special school where there are disabled teachers or teachers who have a communication skill with the disabled.

III. LITERATURE REVIEW

A number of Ethiopian sign languages are employed in varied Ethiopian colleges for the deaf since 1971, and at the first level since 1956. Ethiopian language, presumptively a national customary, is employed in primary, secondary, and at national capital University tertiary education, and on national tv. The Ethiopian Deaf Community uses the language as a marker of identity [4].

Since linguistic communication (SL) may be a visual-spatial language supported point and visual parts, like the form of fingers and hands, the placement and orientation of the hands, arm, and body movements and hand relating

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

totally different elements of the body. The descriptive linguistics structure of Sign language typically has five components. every gesture in SL may be a combination of 5 building blocks. These 5 blocks represent the dear components of SL and may be exploited by machine-controlled intelligent systems SLR.

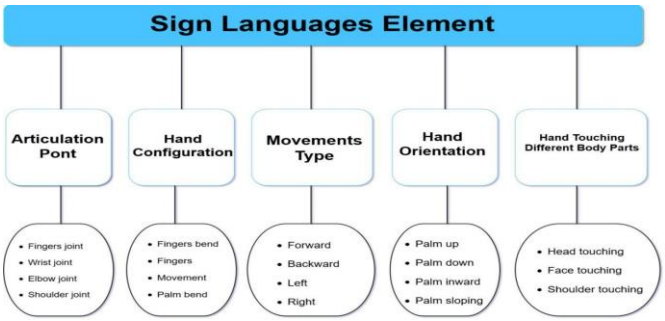

Figure 1 The Ethiopian sign language elements

Figure 2 Some Ethiopian Sign Language Amharic characters

Scholarly interventions to beat disability-related difficulties are multiple and systematic and vary in step with the context. 3 approaches particularly, vision-based, sensor-based, and a mixture of the 2 (hybrid) area unit adopted to capture hand configurations and acknowledge the corresponding meanings of gestures [5].

IV. SYSTEM DESCRIPTION

The system uses a hybrid approach incorporating both sensor-based and vision-based approach to translate a sign language to audio.

A. Vision-Based Approach

A vision-based system or platform is a real-time vision-based system for recognizing Ethiopian Sign Language using a single camera to track the user's signing hands. The system uses Machine Learning Neural Networks Single Shot MultiBox Detector (SSD) on TensorFlow to detect the hand gestures. Since the system uses ML Neural network backgrounds, lighting, skin tone won't affect the detection and gesture recognition. The system generally has three main parts of the approach to recognize Ethiopian Sign Language.

- Hand Detection
- Face detection
- Overlap detection
- Hand gesture and pose detection.

The system receives the video from the webcam in real- time then analyze the video and find out if there are a hand and a face present in the frame. This is to find out if there is a person in the video frame who is about to sign. If the hand and face are detected in the video frame then the system checks for an overlap between the hand and the face of the subject in the video feed. Then identifies the part of the face (eye, forehead, chin, and chick) that the hand overlaps on and identifies what that mean in Ethiopian Sign Language, For example, the hand detected area overlapped with the face detected area and from the face part, that hand detected area overlaps with the chin it gives the definition of 'Mother'; it is shown in Figure 3.

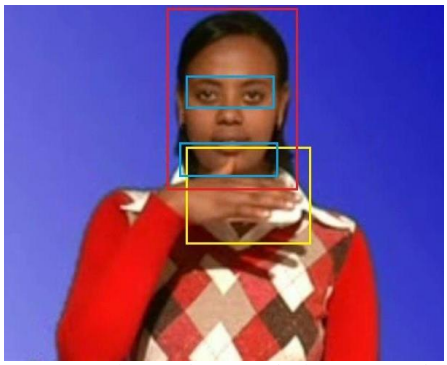

Figure 3 Detection overlap example (it means 'Mother')

B. Sensor-Based Approach

Five flex sensors were attached on the thumb, index, middle, ring and pinky fingers and the palm of a glove in order to measure the bent of the fingers, then the MPU6050 which consist of both the accelerometer and gyroscope and has a six-axis locator was placed on the back of the hand in order to determine the position and movement of the hand on the space. And to connect and communicate the system with the computer we used a Bluetooth module that will enable Bluetooth communication between a computer and the Arduino Uno.

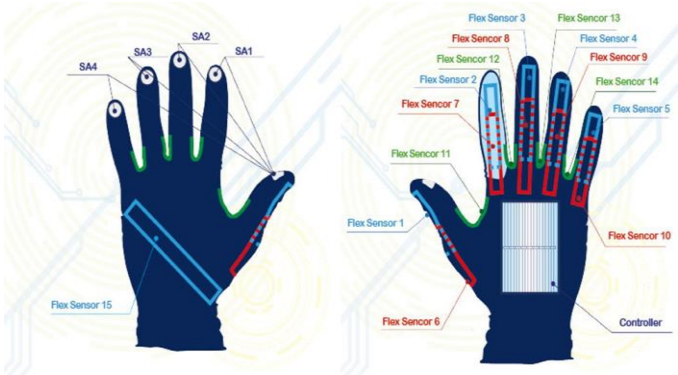

Figure 4 Sensor equipped Glove used in the sensor-based approach

C. Dataset Preparation

Due to unavailability of the datasets for training the machine with our sensors data the dataset for training the model was made where data at the serial port for each gestures representing the alphabets and some frequently used words were collected by using our own developed hardware system and a software called PLX-DAQ. PLX-DAQ is a software receives the data value of the flex sensors and the MPU-6050

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

(Gyroscope 3-axis and Accelerometer 3-axis) and saves the data received in an excel document in a column of their perspective variable in a CSV format.

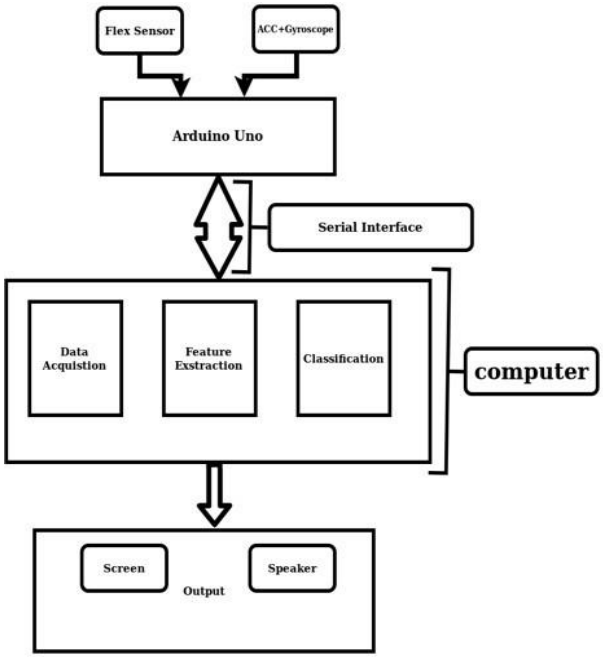

Figure 5 Sensor-based approach block diagram

V. OVERALL SYSTEM ARCHITECTURE

The overall system design has three main components

A. Hand Detection

The first step of the system would be to detect the hand of the signer from the top of the wrist. For this hand detection, we have used Real-time Hand-Detection using Neural Networks (SSD) on Tensor flow. With sufficiently massive datasets, neural networks give chance to train models that perform well and address challenges of existing object tracking/detection. algorithms despite poor lighting, various viewpoints, diverse background, skin tone on the signer and different elements that would distract a hard-coded algorithm OpenCV vision-based detection system.

In hand detection phase when the palm point is found, it will draw a circle with the palm point because of the center point within the palm. The circle is termed the inner circle as a result of it's within the palm. The radius of the circle step by step will increase until it reaches the sting of the palm. that's the radius of the circle stops increasing once the black pixels square measure enclosed within the circle. The circle is the inner circle of the largest radius that is drawn because of the circle with the red color in Figure 6.

When the radius of the maximal circle is acquired, a bigger circle the radius of that is 1.2 times of that of the maximal circle is created. The circle is drawn as the blue color circle within the figure below. Then, some points (X, Y) are sampled uniformly on the circle.

For each sampled point on the circle, its nearest boundary point is found and lined to that. The boundary point is judged in a very easy method. If the eight neighbors of a picture element encompass white and black pixels, it's labeled as a boundary purpose. All of the nearest boundary points found area unit coupled to yield the palm mask which will be accustomed section fingers and so the palm. the maneuver for searching the palm mask is described in rule shown below. a much bigger circle instead of the skin pack is used so on yield plenty of correct palm mask for the next segmentation.

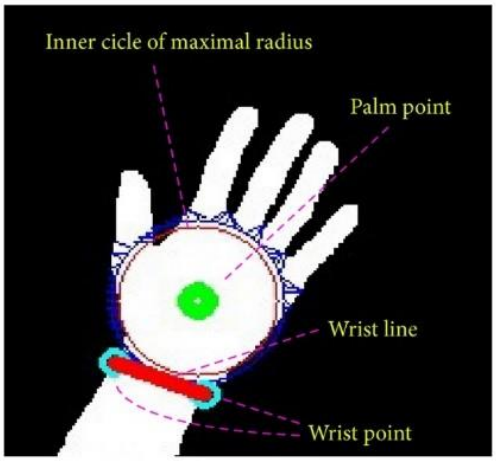

Figure 6 Wrist point, wrist line, wrist point, inner circle, and maximal radius.

Two wrist points are the 2 ending points of the carpus line across the bottom of the hand. The carpus points square measure important points for hand gesture recognition. they'll be searched within the following manner: if the gap between 2 ordered mask points Pi, Pi+1 is large, these two mask points are judged as the wrist points. That is,

$\arg P_i,P_{(i+1)} \max dist(P_i,P_{(i+1)}),P_i,P_{(i+1)} \in S$

Where S is the set of palm mask points and

$dist(P_i,P_{(i+1)})$ is the distance between two points.

Hence, the 8-neighbors of pixel p, denoted by N8(p), include the four-neighbors and four pixels along the diagonal direction located at (x-1,y-1)(northwest),(x-1,y-1)(northeast), (x+1,y- 1)(northwest),(x-1,y+1)(southwest) and (x+1,y+1)(southeast).

Extracting background based on texture and boundary options, identifying between hands and background using color histograms and HOG classifiers, etc. creating them not very sturdy and efficient. that's said to be the most drawbacks to usage for time period tracking/detection is that they will be complicated, square measure comparatively slow compared to tracking-only algorithms and it is quite high-priced to assemble a decent dataset. however, things are dynamic with advances in quick neural networks and totally different completely different / open source free datasets developed by different massive corporations.

This entire hand detection has been created additional approachable by deep learning frameworks like the tensor flow object detection API that change the method of training a model for custom object detection, in our case hand detection. additional significantly, the appearance of quick neural network models like SSD makes neural networks an attractive candidate for period detection (hand tracking) applications.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

Training a model to own associate efficient and satisfying result, maybe a multi-stage method. assembling dataset, cleaning, splitting into training associated take a look at partitions and generating an illation graph. Like any CNN based mostly task, the foremost expensive and riskiest a part of the method must do with finding or making the proper and annotated dataset as a result of the potency of the entire hand detection system hardly rely upon the dataset we choose to train the system. And there are 2 open-source obvious datasets on the online,

- Oxford Hands Dataset and
- EgoHands Datasets

But by comparing these two Datasets from different perspective we have chosen the EgoHands Datasets. It was a much better fit for our requirements for the platform. This dataset is chosen for several reason and it also works well for several reasons.

- It contains high quality, pixel-level annotations (>15000 ground truth labels) where hands are located across 4800 images,
- All images are captured from an egocentric view (Google glass) across 48 different environments; indoor and outdoor and activities like playing cards, chess, solving puzzles, etc.
- There is a possibility to semantically distinguish left and right hands.

Figure 7 EgoHands Dataset contains 48 different videos of egocentric interactions with pixel-level ground-truth annotations for 4,800 frames and more than 15,000 hands.

All the data collected with various features(sensor value) as per the gesture were categorized with their respective alphabet or word as their target value, and the final 21 Feature Extraction collected data were randomized or shuffled in order to reduce the variance and to make sure that the model remains general and overfit less. A total of 15050 data (after all preprocessing steps) taken at different time, condition, setting. Different time includes various timing for data set collection and condition and setting include different position of hand on space for the same category as well as various level of bent of the fingers showing the same category.

B. Face Detection

Detect the face area of the signer using OpenCV object detection system using haar cascade.

C. Hand Gesture Recognition (landmarks and pose of the hand Recognition)

Recognizing landmarks and pose of the hand

For recognizing the landmarks of the hand, we choose to use ColorHandPose3D. ColorHandPose3D is a Convolutional Neural Network (CNN) estimating 3D Hand Pose from a single RGB Image This particular model segments the location of hand present in the picture and gives 21 various landmarks present in hand in 2D and 3D Cartesian Coordinates.

The overall approach consists of three deep networks that cover important subtasks on the way to the 3D pose.

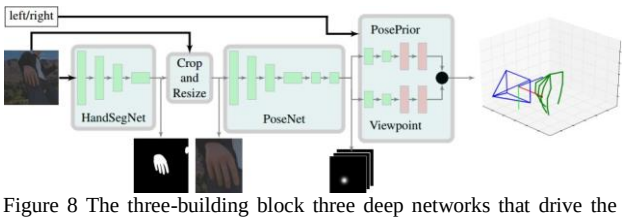

Figure 8 The three-building block three deep networks that drive the 3D hand pose

- Hand Pose Estimation

The first network provides a hand segmentation to localize the hand in the image. The hand is localized within the image by a segmentation network called HandSegNet. And creates the hand mask.

- Localize key-points in 2D

Based on its output, the second network localizes hand key- points in the 2D images. Accordingly, to the hand mask, the input image is cropped and serves as input to the PoseNet. This localizes a set of hand key-points represented as score maps.

- Derive the 3D model

The third network finally derives the 3D hand pose from the 2D key-points. The PosePrior network estimates the most likely 3D structure conditioned on the score maps.

D. Operational Steps

The experimental set-up for gesture recognition is shown in Figure 9 and the steps involved in sign language to speech conversion are described as follows:

Step 1: Take the video of the signer using webcam
Step 2: Pass the video to the system simulation
Step 3: Detect the position of the hand and the face of the signer
Step 4: check for the hand detected area and face detected area overlap
Step 5: If the hand and face overlap, identify the word (term) associated with it else check hand segmentation, map key-point score, map 3D hand pose & hand mark derived, then identify the pose of the hand and derive the term or character associated with it
Step 6: Translate the processed data to speech or text format
Step 7: Display the translated speech or text

VI. RESULT AND ANALYSIS

Currently, the platform can detect Amharic character from Ethiopian Sign Language signer by using hand gesture recognition. And also some word that does have motions by using the detected area overlap approach. So the result shows the platform currently works on signs

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

that do not have motions just a static signs in Ethiopian Sign language. Those are the first Amharic characters and some words.

A. Implementation of Vision-Based approach

The vision-based approach on sign language recognition and translation system works by capturing the signer's hand and analysis the captured gesture using machine learning and deep learning.

The result of Hand detection and gesture recognition is not present here. Since theses recognitions are based on ML and SSD/CNN Neural Network they need some more training on the data and because our processor is slower than ordinary processors that are used in Neural Network. It takes much time one our pc processor.

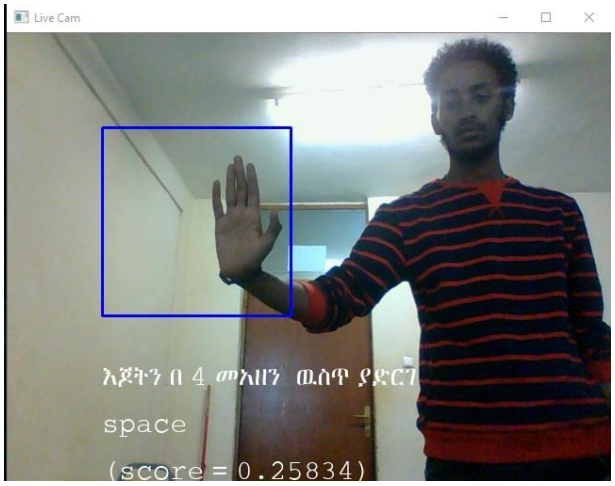

Figure 9 Vison Based Detection

B. Implementation of Sensor Based Approach

The sensor based approach works using different sensors like flex sensor, MPU- 6050 (Accelerometer and gyroscope) that are embedded on a glove to recognize a gesture.

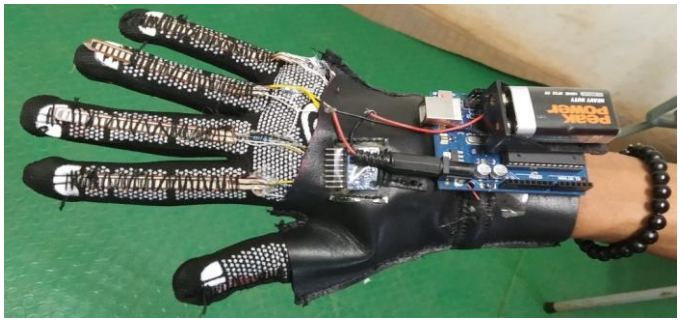

Figure 10 Figure 8 Setup for gesture recognition

Table I Flex sensor voltage value and their corresponding Amharic alphabet

Index value	Middle value	Ring value	Pinky value	Angle (Degrees)	Recognized alphabet
>120	>1	1-120	>170	90	ሀ
150	120	>300	1	180	ፐ
<150	<150	>90	<60	90	ረ
5	>210	>170	1	30	ጨ

Table I shows the voltage values combination for displaying the alphabets ሀ, ፐ, ረ, and ጨ. Likewise different combinations of voltage values are used to recognize the remaining alphabets. Table II shows the voltage values for which the corresponding Amharic word is displayed.

Table II Flex sensor voltage value and their corresponding Amharic words

Index value	Middle value	Ring value	Pinky value	Recognized word	Meaning in English
100	255	>250	1-100	ይሄ	This
>100	>150	255	<100	የኢትዮጵያ	Ethiopian
1	<230	<100	255	የምልክ	Sign
255	>50	1	>50	ቋንቋ	Language

Figure 10 from left to right shows the word "የምልክት" and the word "ቋንቋ" output as audio on the speaker when it is gestured by the user. Similarly the remaining word can be gestured and the corresponding audio is listened.

Figure 10 Gesture L-R "የምልክት" and "ቋንቋ"

Table III Performance Analysis of Sensor Based Approach

	ሀ	ለ	ሐ	መ	ረ	ሰ	ሸ	Performance(%)
1	✓	✓	✗	✓	✓	✓	✓	85.714
2	✓	✗	✓	✓	✓	✓	✗	71.4
3	✓	✓	✓	✓	✓	✓	✓	100
4	✗	✓	✓	✓	✓	✓	✓	85.714
5	✓	✓	✓	✓	✓	✓	✓	100
6	✓	✓	✓	✓	✓	✓	✓	100
7	✓	✗	✓	✗	✓	✓	✓	71.4
8	✓	✓	✓	✓	✗	✓	✓	85.714
9	✓	✓	✓	✓	✓	✗	✓	85.714
10	✓	✓	✓	✓	✓	✓	✓	100

Table III shows the performance analysis of the sensor-based system for gesture recognition. The performance of the system is calculated based on the likelihood of properly recognizing the gesture by the system. within the current work the system is trained for a group of seven gestures. supported user interaction with the setup, the experiment is dispensed for ten times and also the average recognition rate obtained is 88.56% across all

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 08, August-2019

the gestures and for each of those gestures the typical recognition rate is found to be 80-90%.

VII. FUTURE WORK AND CONCLUSION

A. Conclusion

Communication is a base for mankind to live together we are called social animals for a reason. Efficient communication helps strengthen our relationship and expressive intuition in many ways. But when we see the deaf part of our society we can see that how much part of our society was missing from society in many ways. This problem has been around us so long that we know it's there but we chose not to see it or talk about is, it's like an elephant in the room that no one talks about. Considering the communication gap created among the deaf and the hearing that has built a communication barrier wall in the society a platform like this can decrease the problem drastically.

This research has designed a real-time Ethiopian Sign Language translator using a hybrid approach (vision-based approach and Sensor-based) supported by ML and SSD Neural network. The result obtained from the testing of this platform is, the system can efficiently detect, recognize and translates Amharic characters and motion-less words that are signed in Ethiopian Sign Language.

When this platform is fully developed and becomes as efficient as possible, successful implementation of this platform solves a lot of problems in society. There will be inclusive sectors for all the hearing and deaf to use and enjoy them fully. There will be a better education opportunity for deaf students, in hospitals, city administrations, shops and etc. Public service providers (like hospitals, police stations, city administration, etc.), malls, hotels, airports should implement this platform and try to use this system, it's an advantage from different angles like availability to all parts of the society, increasing the number of customers, increasing customer satisfaction and etc.

B. Future work

In future work the system will:

- Detect words and characters that have motions.
- Have an efficient, user-friendly and accessible user interface
- Have a better accuracy because we will get time to train the system a lot more than we have now
- Have a much faster detection and recognition
- Have a lot more coverage of the basic everyday use words
- Work on different platforms in different environments
- Include hardware implementation like proximity and motions sensor for movement estimation of the hand and etc.

REFERENCES

[1] "Country profile study on persons with disabilities," Wa'el International Business and Development Consultant, Ethiopia, 2014.

[2] Tokuda, K.; Nankaku, Y.; Toda, T.; Zen, H.; Yamagishi, J.; Oura, K., "Speech Synthesis Based on Hidden Markov Models," in Proceedings of the IEEE , vol.101, no.5, pp.1234-1252, May 2013.

[3] Ethnologue, "www.Ethnologue.com," Ethnologue, 2013. [Online]. Available: www.Ethnologue.com/subgroups/sign-language. [Accessed 25 December 2018].

[4] W. H. O. D. a. H. Loss, "https://www.who.int/," World Health Organization,[Online].Available: http://www.who.int/mediacentre/factsheets/fs300/en/#content. [Accessed 13 November 2017].

[5] "Country Profile Study on Persons with Disablitites," Wa'el International Business and Development, 2000.

[6] E. Hailu, "Regular BA program in Ethiopian Sign Language and Deaf," SIGN LANGUAGE NEWS AT ADDIS ABABA UNIVERSITY, p. 1.

[7] A. Colgan, "http://blog.leapmotion.com," Leap motion, 9 August 2014 . [Online]. Available: http://blog.leapmotion.com/hardware-to- software-how-does-the-leap-motion-controller-work/. [Accessed 20 January 2019].

[8] g. ilango, "https://gogul09.github.io," 6 April 2017. [Online]. Available:https://gogul09.github.io/software/hand-gesture-recognition-p1. [Accessed 6 January 2019].

[9] J.-T. K. J. L. J. Z. a. Y.-B. Y. Zhi-hua Chen, "Real-Time Hand Gesture Recognition Using Finger Segmentation," The Scientific World Journal, vol. Volume 2014, p. 9 pages, 2014.

[10] E. S. Team, "https://www.expertsystem.com," Expert System, 2017. [Online]. Available: https://www.expertsystem.com/machine- learning-definition/. [Accessed 2 January 2019].

[11] J. Hui, "https://medium.com," Medium, 13 Mar 13, 2018. [Online]. Available:https://medium.com/@jonathan_hui/ssd-object-detection- single-shot-multibox-detector-for-real-time-processing- 9bd8deac0e06. [Accessed 10 January 2019].

[12] R. Kumar, N. K. Chilamkurti and B. Soh, "A Comparative Study of Different Sensors for Smart Car Park Management," in The 2007 International Conference on Intelligent Pervasive Computing, Jeju City, South Korea , 2007.