Sri Lankan Sign Language to Text-Speech Conversion and Vise-versa – EasyTalk

D. Manoj Kumar
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it17050272@my.sliit.lk

K. Bavanraj
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it17032766@my.sliit.lk

S. Thavananthan
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo Sri Lanka
it17068192@my.sliit.lk

G.M.A.S. Bastiansz
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
it17143950@my.sliit.lk

S.M.B. Harshanath
Department of Software Engineering
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
harshanath.s@sliit.lk

Jesuthasan Alosius
Department of Information Technology
Sri Lanka Institute of Information Technology
Colombo, Sri Lanka
jesuthasan.a@sliit.lk

Abstract— **Sign language is used by the hearing-impaired and mute community to communicate with each other. Not all of us are aware of the sign language and we do require a translation. Most of the Sri Lankan Sign Language is tightly bound to the hearing-impaired and mute which makes it difficult for the verbally challenged to communicate with normal people in Sri Lanka. In this paper, we present a translator which will translate the Sri Lankan Sign Language into text-speech and vice versa which would benefit the verbally challenged to express their ideas back and forth.**

Keywords— machine learning, image processing, low-resolution image recognition, convolutional neural networks, natural language processing, real-time translation, semantic analysis, text to speech conversion.

I. INTRODUCTION

Communication is one of the most important roles in the day-to-day life of each living beings. McFarland once said that communication is "*a process of meaningful interaction among human beings. More specifically, it is the process by which meanings are perceived and understandings are reached among human beings*." Every creature on the planet follows their ways to communicate with each other. More importantly, communication brings people together, closer to each other & act as a bridge between individuals & groups through the flow of information & understanding between them. Human beings always use words to convey information to one person to another. But communication methods get varied according to the types of human groups. The best example is the people who can't communicate verbally, who is commonly known as hearing-impaired & mute community. They can't use normal communication methods to express their feelings with ordinary people. This special community use sign language for communication purposes. But unfortunately, ordinary people are unaware of sign language as same as this special community unaware of verbal language. Therefore, building a good communication bridge between the ordinary people & the people with hearing loss or disability to speak seems a bit challengeable task in present days.

More than 250-300 million people around the world are people with hearing loss & speaking disabilities. As mentioned in the above paragraph, sign language had introduced to those people to communicate with each other. Sign Language is a type of non-verbal communication done with body parts, hand shapes, positions & movements of the hand, arms, facial expressions & movements of the lips & used instead of oral communication. Sign language is not universal & they are not mutually intelligible with each other, although there are also striking similarities sign languages. Almost all the countries have their sign language such as Sri Lankan Sign Language[1], American Sign Language[2], British Sign Language[3], Indian Sign Language[5] & French Sign Language[4]. Even these sign languages are unique to each country, ordinary people who belong to those countries are unable to understand them. Because of this, people with hearing & speaking disabilities can't communicate with ordinary people. Therefore the researchers in those countries developed sign language translators which can be used by both ordinary people & people hearing-speaking disabilities while communicating with each other. Hence these translators act as a bridge between these two types of people, they can express their feelings, thoughts with each other very easily. But once a translator introduced it cannot be used by other countries other than the country which introduced it, as sign languages are unique to each country. Because of this reason, nowadays most countries are planning to build translators for their countries.

As assistance for hearing-impaired & mute community, Sri Lanka also has a unique sign language which is also known as Sri Lankan Sign Language (SSL). But problems arise when there is no proper understanding of the SSL among the ordinary people. Therefore, both parties are refusing to be friendly & communicate with each other as they could not understand what each other tries to say. But, to overcome from this problem & to build a communication bridge between ordinary people & people with difficulties with speaking & hearing Sri Lanka also invented a few translators. Unfortunately, most of the invented translators are based on activities in hearing-impaired & mute schools, therefore, ordinary people unable to get the maximum use of them. On the other hand, they are capable of translating SSL into natural vocal language or translating natural local language into SSL, it is hard to find out a system which can do both translation modes using one interface.

From this research, we are planning to come up with a translator which can translate SSL to text/voice & vice-versa.

The rest of the paper will describe the research background, the methodology used, results and discussions, conclusion and the future works of the authors regarding this new translator.

II. Background & Literature Survey

Researches related to sign languages are done around the world. Chinese sign language [8], American sign language [9], Bangla sign language [10], Marathi [11] and Sinhala [12] are some of them. These sign language detections have two approaches. One is image or vision-based and the other one is device-based [13][14]. In general, the identification of signs includes pre-processing, extraction of features, and classification in a simple phase. Pre-processing is done in the sense to remove an unwanted element from the context. From the Region of Interest (ROI), image feature will be extracted and it will be classified using a certain method [14][15].

When talking about device-based sign recognition, a system is proposed by Gibran & team for Mexican sign language using Kinect sensor [16]. The system will store the colour, depth and the skeleton tracking information using the RGB-D camera. Dynamic Time Wrapping (DTW) algorithm had been used in this system to interpret gestures. For testing, they have used the K-Fold Cross-validation approach. This check demonstrated a 98.57% mean accuracy in real-time testing. Similarly, Anant Agarwal & team have suggested a program that will use Microsoft Kinect [17]. As a classifier, they used the Support Vector Machine (SVM) algorithm and compared their results with existing techniques. Their work has proven that SVM with Radial Basis Function (RBF) kernel can give more classification accuracy than the linear kernel. Deepali's team has proposed a system using leap motion controller which gave them 96% accuracy by using a dataset with 520 samples [19].

Haar Cascade Classifier has shown 92.68% accuracy in Kanchan & Surekha's work [18]. They have interpreted Indian sign language using webcam images. However, some images are not classified because of high motion gestures. For the same sign language, Probabilistic neural network (PNN) classifier has shown more accuracy than K-Nearest Neighbor (KNN) classifier [14].

III. Methodology

The product comprises of four major components namely: "Hand Gesture Detection", "Sign Recognition & Translation", "Text & Voice Assistance" & "Text to SSL Conversion". Basically, the system is created to convert common words which are used in our day-to-day lives such as "Good Morning", "How are you?" etc. To fulfil the system's objective, the system database consists of 2 main tables where one table with signs for pre-defined common words & the other table with signs relevant to the English Alphabet.

A. *Hand Gesture Detection*

This component will be done on top of TensorFlow models and using Faster RCNN configuration. For our model training, 03 images per letter in the Sri Lankan Sign Language Alphabet (English) [1] total of 247 images. First, a label map is created for the classes that are involved. In this scenario, there is only one class: *hand.*

The images will be taken in 800 x 600 resolution to facilitate faster training and storage purposes. All the images will be taken using low-resolution laptop webcams. The reason is to get low-resolution images as the base of the training process. So when our model runs, it will be detected even the low-resolution images. Also, this will be an advantage in low latency network connections where there is no much network traffic (uplink & downlink) is necessary. This is to facilitate the people in rural areas of Sri Lanka[21] who have low-speed internet connections.[20]

Then split images into train and test sets and label the images using labelling tool in Pascal VOC format and get an XML for each image. Then convert those individuals into CSVs for train and test image sets. Using the CSV files, start training the model using TensorFlow model and Faster RCNN Configuration. A 30min-1 hour training would be sufficient and check for the loss in the terminal and stop it once it reaches below 0.25.

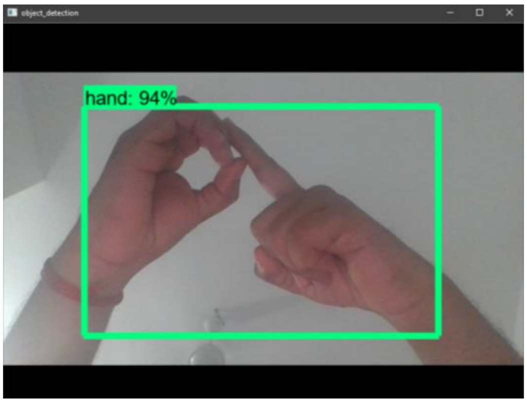

Fig. 1. Detecting Hand Signs

After the successful model training, the webcam opens and detects the images using the pre-trained model. Once it detects an image with an accuracy of more than 95%, the image will be sent as a POST request to the next component's (Image Classification) API.

B. *Image Classification*

The image classifier is wrapped under an API where it gets the abstracted sign image as the input and predicts the English alphabet letter. The API is a machine learning model which is built using convolutional neural networks (CNN). Convolutional networks were developed using the concept of neuron connectivity pattern in animals and it is widely used in visual analyzing. CNN classifier can process an image and classify it under different categories. Since we are using python we chose "keras" library which is pretty easy to build a CNN model. When training a CNN model using a large number of data will help to predict more accurately. In our context, we used a dataset of English sign alphabet letters. Every letter of the alphabet is a class. So, we have 26 classes, each with 3000 hand sign pictures, which are taken from a low-resolution camera.

An image in CNN classifier will go through several steps. As the first step, the image will travel through a sequence of convolution layers with kernels. Dimensionality reduction will happen to the image to highlight the important data in it. This step is called pooling. Afterwards, many convolutional layers will be added, and pooling will happen repeatedly until the filtration is satisfied. Then the output is flattened and passed through a fully connected layer. Finally, the class will be identified by applying an activation function. The signed letter we predicted using CNN will be sent through the API to the text, voice assistant.

C. *Text & Voice Assistant*

Identification of the words segments and conversion of the text into word segments requires some understanding of the computer language. I usually use the Linguistic Data Consortium. I use the data texts in a collection of texts known as corpus which is a sequence of words and pronunciation known as tokens. In this regard, we aim to create a method of solving various problems in data sets and tokens.

Define a probabilistic model: This is the incorporation of random variables with a probability distribution of the model of an event and the data collected is used in the determination of the probability of each of the candidates.

Enumerate candidates: This is an analysis of the candidates in the race by listing them and showing their performance in the field.

Choose the most probable candidate: This involves the selection of the best candidate by the use of the language model to get one with a high probability.

This can be done by mathematical equations as shown:

$$best \ = argmaxc \ \in \ candidates \, P(c)$$

Or, you can also use Python in computer code that involves the use of the following formula;

$$best \ = \ max(candidates, key = P)$$

The probabilistic language model is the probability that involves consideration of a sequence of words. This is usually emphasized on the probabilities of each of the words that are analyzes in the context of the word in all preceding words.

In equations for instance

$$P(W1{:}n) \ = \ \Pi k = 1{:}nP(Wk \mid W1{:}k{-}\,1)$$

This can be applied in the segmentation, in which we define a segment or a function that acts as the input of a string of a list of words that have no spaces with the best segmentation possible

>>> segment('choosesrilanka')
['choose', 'Srilanka']

In spelling checking with an input word, the segmentation method can be used to determine what the initial intention was. The use of Bayes' theorem becomes effective at such instances which aid in the precise determination of the right candidate;

$$P(c \mid w) \ = \ argmaxc \, P(w \mid c) \, P(c)$$

Whereby P(c), the model language is represented by the probability of c being the intended word, while P(w | c), represents the probability that the error model in which the author intended to write c but wrote.

In Table I below there is an indication of the words w and c and the various probabilities of w and c and their products.

TABLE I.
The product of the probabilities

W	c	w | c	P(w|c)	P(c)	10^9 P(w|c)P(c)
thew	the	ew | e	.000007	.02	144
thew	thew		.95	.00000009	90
thew	thaw	e | a	.001	.0000007	0.7
thew	threw	h | hr	.000008	.000004	0.03
thew	thwe	ew | we	.000003	.00000004	0.0001

This stage involves the substitution of the words that have been found fit and matching in the English corpus. The impact in this regard is that there is also a loss of speech and voice of words substituted in the output.

The process involves two sections; natural language processing and speech digital processing. NLP involves the interaction between humans and computers by the use of natural languages. This is done by the use of the prosodic feature in the input system that involves three components in text analysis, prosodic phrasing, and phonetic conversion. Text analysis, for instance, is the segmentation of the initial input sentence into tokens in which each of the words is analyzed as part of the speech (POS). The phonetic conversion dictionary is used as an approach for transcription of phonetic in the initial input words. In this case for the input text word to run it must be included in the dictionary. Finally, the prosodic phrasing method involves the classification of the functions and the content word accompanied with some modifications.

Finally, speech synthesis is the production of speech in the most natural way possible with intelligible sounds. Concatenative is the most natural method as compared to the rest of the methods. The method involves the selection of optimum sets of acoustic units from the speech database to have a match with the phoneme stream and the targeted prosody. Phoneme entails concatenation of the phonetic units informing of the word while domain-specific uses concatenate of prerecorded words in completing the required utterances.

D. *Text to SSL Conversion*

Text to SSL Conversion component can be used as an SSL learning material for ordinary people. Through this component, users can get the relevant hand signs for verbal texts. Converting text into SSL in real-time is done with the support of several libraries in Natural Language Processing (NLP)[7] & semantic analysis[6]. More importantly, NLP is used to understand, analyze, manipulate, potentially generate human language & Semantic Analysis is used to create GIF images using selected hand signs. The system overview diagram for this component is shown in below Fig. 2. figure.

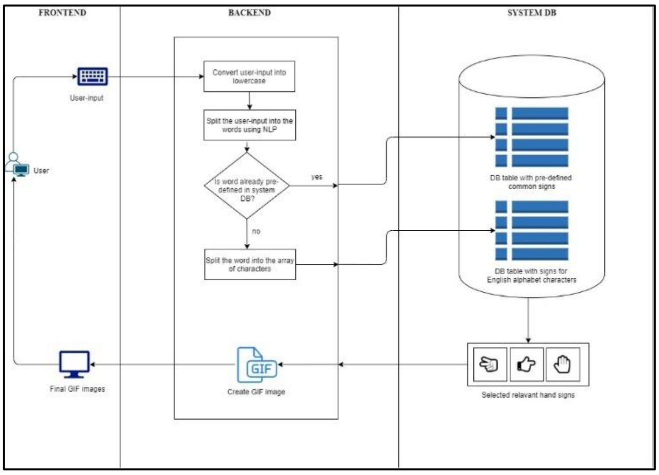
FRONTEND
BACKEND
SYSTEM DB
User-input
User
Convert user-input into lowercase
Split the user-input into the words using NLP
Is word already pre-defined in system DB?
yes
no
Split the word into the array of characters
DB table with pre-defined common signs
DB table with signs for English alphabet characters
Selected relavant hand signs
Create GIF image
GIF
Final GIF images

Fig. 2. System Overview Diagram

The user input will go through several steps before the system displays a GIF image of the sign as the final output. Once the user inserts the input, it will convert into the lowercases in the system backend. Then the system will split the input into multiple words using nltk library in NLP. Afterwards, the split words will compare with a set of pre-defined common words to identify whether there is any similarity between user-input & pre-defined words. If there is any similar word, then the system will get the relevant hand sign from the system database & resend back to the backend. Otherwise, the word will split into an array of characters again & pass that array to the system database. Once selects all the relevant signs for the characters, the system will order them according to the order the characters & send it back to the backend. After the previous step fulfils, by using Semantic Analysis the system will create separate GIF images for each word in the user input & displays to the user as the final outcome.

IV. Results & Discussion

This section elaborates the experiment results conducted on each model. Maintaining a better communication experience between ordinary people & mute, hearing-impaired people in Sri Lanka by introducing a Sri Lankan sign language translator which can translate SSL into text/audio & vice-versa in the main objective of this research. High accuracy must be fulfilled to ensure accurate translation while reducing the loopholes which can happen during the communication. A simple survey was conducted to identify different problems arise while communicating with verbally impaired, hearing-impaired people and to identify the ideas of the Sri Lankans about the translators. The results of this survey specify multiple main features. First, most of the people do not have a clear picture of the sign language translators. Also, they don’t have enough interest to learn sign languages because no one in the society spoke about the importance of learning sign language. Secondly, in Sri Lanka, there is no proper communication system between ordinary people & verbally impaired, hearing-impaired people. The main reason for this feature is the same, not having a proper impact to learn sign language. The solution was designed targeting the needs of the people & considering the Sri Lankan sign language. The most important feature of this application is hand gesture detection was trained using faster R-CNN based model.

At the beginning of product development MASK RCNN based model was selected to train and detect hand sign from the live webcam feed. This was trained using 247 images. The main drawback of this model was it only works for static image input rather than a live feed of images. To avoid this drawback, it was decided to use R-CNN based model which was fast as well as the best-resulting model.

The main fact of Faster R-CNN model is not only classifying the images but also detecting more than one object in an image. This can be identified as the most important factor to detect the hand sign even in the low-resolution laptop web cameras as well. This has experimented with an image that was taken by a low-resolution laptop web camera. Also, the model was tested against the live video through a low-resolution web camera. Faster R-CNN was able to identify the hand gesture in both mentioned scenarios with a higher accuracy level. One of the tested images is displayed in Fig. 1 above.

V. Conclusion & Future Works

In this paper, we present you an application which could translate Sri Lankan Sign Languages to text & voice and vice versa. First, to detect the hand signs, we used R-CNN based model and to translate them, we used the ML-based API which we developed. This API can be used in future developments which are related to the sign languages. Developers do not have to build a classification model from the beginning. They can just use this API with a valid dataset. The text & voice assistant acts as to identify the words segment from collections of alphabets then spelling correction and convert word segment to speak them. We used the NLP based API which we developed. They can use this API for NLP related languages translation not only for sign language. We also reverse-engineered the process for the regular people to get an idea about sign language. So, the text to sign language translator does convert the text sent by the user into GIFs of corresponding Sign Languages using Semantic Analysis.

For the moment, the system is proposed to be a web application and soon will be made into a mobile application with faster responses and lower processing time. Further, with the introduction of 5G, the response times will be faster.

References

[1]"Alphabets,"2007,[Online].Available: http://www.rohanaspecialschool.org/wpcontent/uploads/2010/08/18Alphabets.pdf . [Accessed 13 January 2020].

[2] "American Sign Language," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/American_Sign_Language. [Accessed 13 January 2020].[3] "British Sign Language," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/British_Sign_Language. [Accessed 13 January 2020].

[4] "French Sign Language," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/French_Sign_Language. [Accessed 13 January 2020].

[5] "Indian Sign Language," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Indo-Pakistani_Sign_Language. [Accessed 13 January 2020].

[6] "Natural Language Processing - Semantic Analysis," [Online]. Available: https://www.tutorialspoint.com/natural_language_processing/natural_language_processing_semantic_analysis.htm. [Accessed 02 February 2020].

[7] "NLTK 3.5 documentation," [Online]. Available: https://www.nltk.org/. [Accessed 01 July 2020].

[8] Yaofeng Xue, Shang Gao, Huali Sun, Wei Qin, “A Chinese Sign Language Recognition System Using Leap Motion”, 2017 International Conference on Virtual Reality and Visualization (ICVRV).

[9] Kshitij Bantupalli, Ying Xie, “American Sign Language Recognition using DeepLearning and Computer Vision”, 2018 IEEE International Conference on Big Data (Big Data).

[10] Md Azher Uddin, Shayhan Ameen Chowdhury, “Hand Sign Language Recognition for Bangla Alphabet using Support Vector Machine”, 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET).

[11] Ashish s. Nikam, Aarti G. Ambekar, Bilingual Sign Recognition Using Image Based Hand Gesture Technique for Hearing and Speech Impaired People”, 2016 International Conference on Computing Communication Control and automation (ICCUBEA).

[12] H.C.M. Herath, W.A.L.V.Kumari, W.A.P.B Senevirathne and M.B Dissanayake, “IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE”, SAITM Research Symposium on Engineering Advancements 2013 (SAITM – RSEA 2013).

[13] J. S. Sonkusare, N. B. Chopade, R. Sor, and S. L. Tade, "A Review on Hand Gesture Recognition System," 2015 Int. Conf Comput. Commun. Control Autom. , pp. 790-794,2015.

[14] Umang Patel, Aarti G. Ambekar, “Moment Based Sign Language Recognition For Indian Languages”, 2017 Third International Conference on Computing, Communication, Control And Automation (ICCUBEA).

[15] Kusurnika Krori Dutta, Satheesh Kumar Raju, Anil Kumar G, Sunny Arokia Swarny, “Double Handed Indian Sign Language to Speech and Text”, 2015 Third International Conference on Image Information Processing.

[16] Gibran García-Bautista, Felipe Trujillo-Romero, Santiago Omar CaballeroMorales, “Mexican Sign Language Recognition Using Kinect and Data Time Warping Algorithm”, Available at: ieeeexplore.ieee.org. [Accessed 13 Feb. 2020].

[17] A. Agarwal and M. K. Thakur, "Sign Language Recognition using Microsoft Kinect", 2013 Sixth International Conference on Contemporary Computing (IC3).

[18] Kanchan Dabre, Surekha Dholay, “Machine Learning Model for Sign Language Interpretation using Webcam Images”, 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA).

[19] Deepali Naglot, Milind Kulkarni, “Real Time Sign Language Recognition using the Leap Motion Controller”, 2016 International Conference on Inventive Computation Technologies (ICICT).

[20] Telecommunications Regulatory Commission of Sri Lanka(TRCSL), Statistics of the month - March 2020. [Online]. Available: http://www.trc.gov.lk/images/pdf/1stQuater2020.pdf. [Accessed 12 June 2020].

[21] UNESCAP, 2013. [Online]. Available: https://www.unescap.org/sites/default/files/SriLanka.pdf. [Accessed 15 May 2020].