TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4890-4894

American Sign Language Translation using Firebase

[1]Shishir K, [2]Spoorti A Tondihal, [3]Rajesh I.S, [4]Srijith S, [5]Shreyas S Rigvedi

[1,2,3,4,5]C&IT, REVA University, Bangalore, India
[1]shishirk2801@gmail.com, [2]spoorthi278@gmail.com, [3]is.rajesh081@gmail.com,
[4]srijiths1998@gmail.com, [5]shreyasrigvedi@hotmail.com

Article Info
Volume 83
Page Number: 4890-4894
Publication Issue:
May-June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Abstract

Sign language is the manual way of communication by the speech impaired people. They use hand gestures and expressions to express their thoughts. Thus it is hard for them to communicate with others. Our project provides a solution by converting the sign language to speech, helping speech impaired people to communicate with others. The Application uses the camera of the device to capture the sign and sends to a model hosted on FIREBASE and the matched sign is converted to speech by GOOGLE'S translator.

Keywords: *American sign language (ASL), Firebase, Speech, Hand gesture, CNN.*

1. Introduction

The only way to express our thoughts, beliefs, emotions and opinions is speech and thus, we regard this aspect of our lives as extremely basic and fundamental. But not everyone is fortunate enough to have this ability. Speech impaired people can only rely on sign language (hand movements and gestures) to communicate. But not everyone can understand sign language. This makes communication extremely complicated for them. Acknowledging this problem, the goal and motive of this paper is to provide a platform which removes the barrier between the able and the disabled. So we are presenting an application that translates the American Sign Language (ASL) into speech which acts as a communication interface for speech impaired people. American Sign Language is a natural language which serves a wide range of basic and widely accepted sign language conventions. We have adopted ASL for the proposed application as it covered all the necessary gestures and would contribute to the accuracy of the system. The application is developed using KOTLIN and is integrated as a mobile phone application. We have used the ML Kit feature of Firebase, which is one of Google's platforms providing many application development features, to process the sign language conventions and translate them accordingly [1].

The rest of the paper is illustrated as follows. Section 2 gives a gist of the previous works in the domain. Section 3 explains the architecture and the components of the system. Section 4 describes the processes and the methodology implemented for the functioning of the application. Section 5 gives the overall outcome of the project and accounts for the improved establishments. Section 6 concludes the paper and throws light on the future latitudes of the venture.

May-June 2020
ISSN: 0193-4120 Page No. 4890-4894

2. Literature Survey

Table 1: Literature Survey for Sign Language Translation

Authors	Method	Accuracy	Remarks
[2]Aarthi M, Vijaylakshmi P	Flex sensors and Atmeg 328 are used for the sign detection.	87.5%.	-Requires glove which is an extra hardware. -Uses very less power and is portable.
[3]ParamaSridevi et al	Image processing using MATLAB.	85.2%.	-Model is implemented in the PC or Laptop.
[4]Keerthi S Warrier, JyateenKumarSahu	-Graphical designing platform called LabVIEW is used. -Image transfer between smartphone and host computer is done by using DroidCam android app and client software	-	-App is not developed; using PC's everytime for communication is impractical.
[5]OmkarVedak, Prasad Zavre et al	-Image processing using combinational algorithm. -Recognition is done using template matching.	88%.	-The dataset used for training is huge, producing remarkable accuracy. -Mobile application has not been built.
[6]Mohammed Elmahgiubi, Mohamed Ennajar et al	-Glove fitted with Flex sensors, contact sensors and 6 DOF accelerometer on a single chip is used	96% (only 20 out of 26 letters).	-using glove, extra hardware. -The model detects only 20 alphabets.
[7]Nikhil Kasukurthi, BrijRokad et al	-Trained the pre-trained model on the Squeeze Net model. The Model is trained on the Surrey Finger. -the image is processed by dividing every pixel by 255 and then resizing the image to 244X244.	83.29%	-The model developed was a squeeze net architecture which enabled the complete architecture to be stored on a mobile
[1]YellapuMaduri et al	-Mobile application is developed using LabVIEW software	-	-The model is incorporated into a mobile phone, hence reducing the hardware. -Translates alphabets and numbers
[8] Malli Mahesh Chandra, S Rajkumar, , Lakshmi Sutha Kumar	Prototype using SVM classifier	-	-Vietnamese Sign Language translation. -Multiclass SVMs.

3. System Architecture

The system architecture for the application can be precisely described with the help of a flowchart sketched in Figure 1.

The user is presented with an interface to interact with the application which forms the first component of the system. The second component of the system is the Firebase Authentication service which provides the authentication feature for the application. The camera of the device holding the application also acts as a component to the application as it captures hand gestures as frames and sends them to the next component of the application which is the Firebase ML kit. The ML kit hosts a model that is responsible for matching the sign with the particular character. The character recognized by the model is presented to the user as the final output by converting it into speech using the Text to Speech method provided by

May-June 2020
ISSN: 0193-4120 Page No. 4890-4894

Google. The responsibilities and the functions carried out by these components can be deduced with the modules identified in the application.

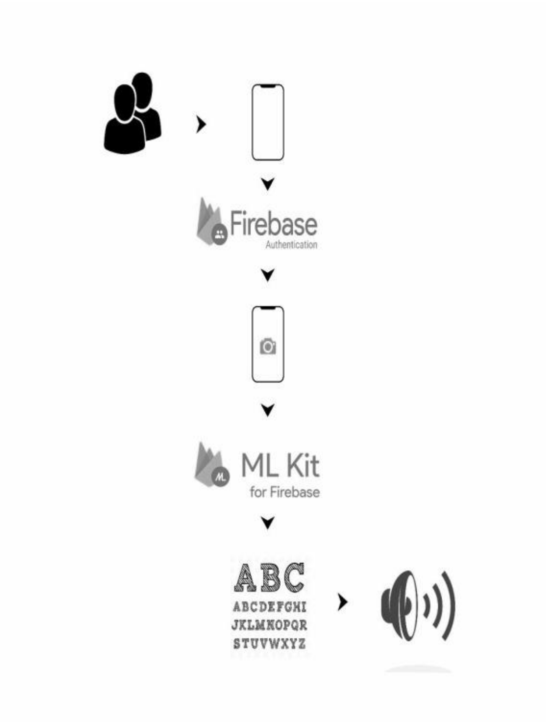

Figure 1: System Architecture

4. Methodology

The modules identified in our methodology are:
Authentication: Authentication plays an important role in recognizing the users using the application. It also helps in logging the actions of the user individually. We have used Firebase authentication as a back end platform for our authentication service adhering its dependable availability, wide variety of tools and ease of use. Recognizing a user's identity enables the application to keep it's matrix secure by providing it's services only to a trusted and acknowledged users. This also makes sure to save every user's data in a cloud environment and thus providing the same personalized experience across all of the user's devices. Google's Firebase provides Authentication as a service among many of it's back end instances. It implements easy-to-use Software Development Kits, and ready-made User Interface libraries to authenticate users of the application. It supports authentication using the generic email and password, phone numbers, popular federated identity providers like Google, Facebook and Twitter, and more. Firebase Authentication integrates tightly with other Firebase services, and it leverages industry standards like OAuth 2.0 and OpenID Connect, so it can be easily integrated with your custom back end.
Image Acquisition: The authenticated user is then provided with an interface where the user can present the gesture through the camera of the device. The basic method implemented in the application to realize the translation involves processing the spontaneous images to extract the constructive features of the image. This makes image acquisition a very important aspect. We make use of a tool called the CameraX to handle our input images. CameraX is a jet pack support library, built to help make the camera application development easier. It provides a consistent and easy-to-use API surface that works across most Android devices, with backward-compatibility to Android 5.0. Rather than storing or sending each image from the frame to the model, we make use of a use case present in the CameraX tool called 'image analysis'. The image analysis use case provides the application with a CPU-accessible image to perform image processing, computer vision, or machine learning inference on. The application implements an analyze method that is run on each frame which is of the type 'Image Proxy'. Each frame is then converted to a Bitmap format after which it is scaled to deduce a Bitmap which is an image with required height and width (64 x 64). This final render serves as the input to our model.
Gesture recognition: The input to the model is the bitmap which is the output of the previous module. The model is hosted on Firebase, using the ML Kit feature. ML Kit comes with a set of ready-to-use APIs for common mobile use cases. This is a custom made model with six layers of Convolutional Neural Network with three dropouts, a flatten and two dense layers.
A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. Using this algorithm, we are able to understand the user's hand gesture properties. Dropout refers to dropping out units (both hidden and visible) in a neural network.

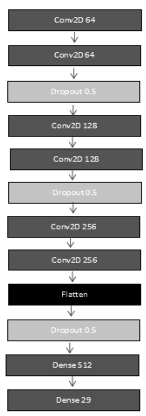

Figure 2: CNN layout

TEST Engineering & Management
May-June 2020
ISSN: 0193-4120 Page No. 4890-4894

Text to Speech: Each letter distinguished by the model is stored in the processor until the complete input object is formed which is identified by a full stop that indicates the termination of the input. The final output that is rendered is formed by integrating all the individual outputs from the model. This formation is then converted to speech format using 'Text To Speech' library presented by Android. 'Text To Speech' as the name suggests , is nothing but the synthesis of artificial human speech fabricated from a text string. The attributes or the features of the speech can be tailored according to the user. Some major characteristics include language, pitch and modulation. There's also a provision to include customized speech inputs through induced files.

5. Results and Discussion

The application is up and running with all essential features and purposes. The User Interface is easy on the eyes and is extremely accessible. Fig 3 shows the authentication stage or the login interface for the application.

Figure 3: Login Interface

The model hosted on Firebase ML kit has a set of 29 classes each representing a particular element in the sample space the user's input. Each class is vigorously trained by the model with elaborate training instances. Each training instance for a class has 3500 images to maximize the accuracy of our model. So the total number of images used to train the model is 101500 images. Validation of our model was performed by randomly processing 29 images representing each character accounted for the input. Fig 4 shows an instance of recognitaion of the alphabet O.

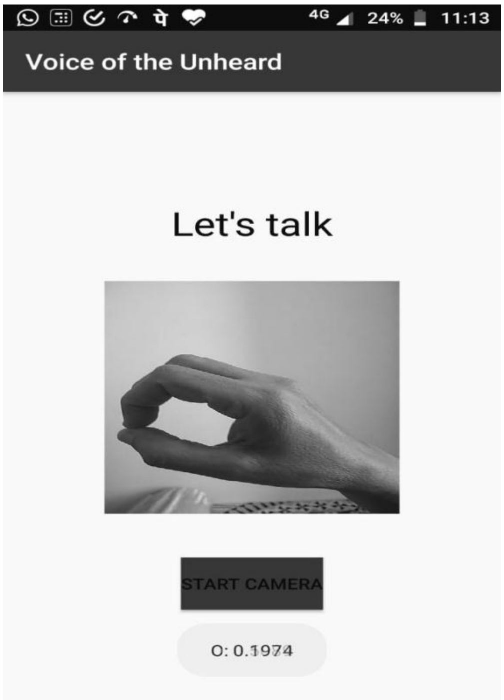

Figure 4: Sign language detection

The proposed idea proved effectively and the application was able to translate the American Sign Language hand gestures into speech effectively. The accuracy of our training data set was observed to be 95.23 percent while the validation set gave an accuracy of 82.94 percent.

Previous Work Drawbacks	Proposed Solution
Gesture Recognition requires gloves.	Sign language images are captured through the camera without the inconvenience of gloves.
Sentence wise translations complex to achieve.	Provision for sentence wise translations with an input termination format.
Services like authentication and machine learning classification had to be performed with manual functions and models.	With the employment of Firebase, Authentication and ML Kit come as it's chunks.

6. Conclusion

Communication may take different forms and can be achieved through various parameters or languages but the substantial need for it throughout the globe makes it an absolute necessitate, irrespective of a person's characteristics. Our project aims to traverse the discontinuity between the enabled and impaired by presenting our application as a bridge. Our approach

May-June 2020
ISSN: 0193-4120 Page No. 4890-4894

is oriented to make our application an acceptable and user friendly one. The idea of a prototype application rather than a mechanized instance like a glove and a method facilitating singular letter or word translations as well as sentence wise translations are some notable characteristics of our project. Thus, it can be concluded that our product surpasses the existing methods and works in the field considering the accuracy and the accessibility.

Future Scope:

We intend to carry out further research in our domain and hope to come up with an enhanced and improved version of our work. The immediate action plan would be to enable the system to provide for communication in both ways i.e. The application, along with translating sign language into speech should also be able to interpret speech into sign language. This would require a reverse engineering technique and would be quite an exciting challenge.

Acknowledgment

We would like to thank our college, REVA University for providing us an opportunity to present our idea as a project.

We would also like to express our gratitude to the Director of our department, Dr. SunilkumarManvi for inspiring us to come up with innovative and deviceful propositions.

We would also like to take this opportunity to thank our guide, Prof. Rajesh I.S who has been there with us in every step and has helped us in every way possible.

Lastly, we would like to extend our thanks to Prof. Manjunath P.C who has supported us and mentored us in every venture of ours.

References

[1] Yellapu Madhuri,Anitha.G, Anburajan.M " Vision-based sign language translation device", Department of Biomedical Engineering, SRM University, kattankulathur, Tamilnadu, India..

[2] Aarthi M., Vijayalakshmi P., "Sign Language to Speech Conversion", Fifth International Conference on Recent Trends in Information Technology (ICRTIT), April 2016.

[3] ParamaSridevi, TahmidaIslam, Urmi Debnath, Noor A Nazia, Rajat Chakraborty, Celia Shahnaz,"Sign Language Recognition for Speech and Hearing Impaired by Image Processing in MATLAB", December 2018.

[4] Keerthi S. Warrier, Jyateen Kumar Sahu, HimadriHalder, RajkumarKoradiya, Varun Raj, "Software based sign language converter", 2016.

[5] Keerthi S. Warrier, Jyateen Kumar Sahu, HimadriHalder, RajkumarKoradiya, Varun Raj, "Software based sign language converter", 2016.

[6] OmkarVedak, Prasad Zavre , AbhijeetTodkar , Manoj Patil, "Sign Language Interpreter using Image Processing and Machine Learning", International Research Journal of Engineering and Technology (IRJET), April 2019.

[7] Nikhil Kasukurthi, BrijRokad, Shiv Bidani, Dr. AjuDennisan, "American Sign Language Alphabet Recognition using Deep Learning", May 2019 .

[8] Malli Mahesh Chandra, S Rajkumar, Lakshmi Sutha Kumar "Sign Languages to Speech Conversion Prototype using the SVM Classifier" TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON).