Universidade Federal de Pernambuco
Centro de Informática

Machine Learning for Hand Gesture Recognition Applied to LIBRAS (Brazilian Sign Language)

Discipline: Introduction to Computing - IF668[X]

Students	Eric Vinícius de Lima[X]
	Michel Leonidas Aleixo da Silva[X]
Professor	Adriano Lorena Inacio de Oliveira[X]
Time	Tuesday - 10:00-12:00 AM

Recife, October 24, 2021

Contents

1   Introduction

After an introductory class on Machine Learning given by Professor Adriano L. I. Oliveira [X], a project was requested with the general objective of applying this knowledge of Artificial Intelligence in practice. Building a program that involved machine learning with models based on Text, Images, Numbers or Sounds, applicable to any real or non-real situation.

The Brazilian Sign Language - LIBRAS - is considered official in the country since 2005 [X], but even with more than 9 million hearing impaired people [X], it is still not taught in public schools, making this language learning even more difficult and, consequently, the communication with a hearing impaired.

With that, we decided to develop an application that uses Machine Learning capable of judging with precision if the gesture made with the hand is or not a vowel of the Brazilian Sign Language, through a webcam.

Figure 1: Inclusion and accessibility of hearing impaired [X]

2   Tools Used

For this experiment, the programming language Python [X] was used together with the Mediapipe[X] and OpenCV[X] frameworks, in addition to the Machine Learning for Kids [X] platform for processing the results with the Watson Assistant from IBM[X].

Figure 2: Tools used in this project

3   Experimental Procedure

Initially, it was studied how the recognition of a hand by the computer works. We can usually represent it computationally as a 21-vertex graph. To reduce complexity, we decided to work with just 11 of them.

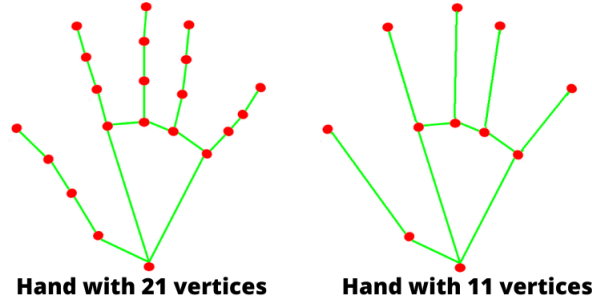**Hand with 21 vertices**   **Hand with 11 vertices**

Figure 3: Hand with different amount of vertices

Also, it was decided to judge only the vowels and not the entire alphabet in LIBRAS. Machine Learning for Kids allows only 10 labels to work with numbers. Each vertex has a coordinate containing 2 values: x and y axis. An impediment was then introduced, so it was decided to further reduce the number of vertices, now to just 5. The most relevant vertices were then chosen, following the following gestures:

A   **E**   **I**   **O**   **U**

Figure 4: Vowels in LIBRAS with main vertices

Following the vertex naming adopted by Mediapipe, the chosen vertices were: 0-wrist, 4-thumb-tip, 5-index-finger-mcp, 12-middle-finger-tip and 20-pinky-tip. With that, the elaboration of the project code was started, and it was divided into 3 parts:

3.1 Data Collect:

This first part of the project was to use the mentioned libraries to collect the data and treat. The OpenCV was then used to recognize a camera and perform some image settings. Then, we use some solutions present in Mediapipe, to detect and recognize hands, in addition to printing them on the screen. MediaPipe treats the hand with 21 vertices, so we treat these vertices to only take 5 of them and eject the z-axis as it is not needed for the project.

Finally, it was possible to collect, through keys pressed on the keyboard, the x and y axes of each of the chosen vertices.

The Input Image   **The Output Data**

Figure 5: Collecting Data

3.2 Populate Training Database:

To communicate with Machine Learning for Kids, we use its API, in which it is possible to request the classification of data and also request the filling of the training database, and integrate it into the project through the Python Request library [X]. For this, another module responsible for making this communication, was created in addition to generating an API access key directly in Machine Learning for Kids, as well as performing other configurations, such as creating labels for the vowels that will be filled.
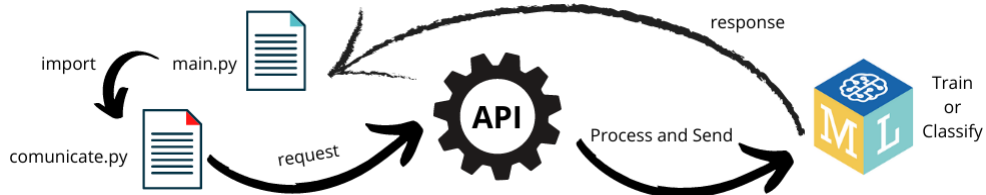

Figure 6: API Diagram

3

3.3   Training the IA:

All computer training is performed by Watson [X], who uses Convolutional Neural Network and linear regression models for classification. Thus, 50 gestures were initially added for each vowel, made by different people, with the right and left hands and on different webcams, to generate greater precision in the results.

Figure 7: Training Diagram

The data entry process for the training database has been automated, now being done by pressing a keyboard key, specific for each vowel gesture. To facilitate the visualization of the main vertices, it was decided to highlight these points in the webcam window. It was also decided to print in this window a text with the return of the request made.

4   Results Analysis

During the analysis of the results, which took place on 8/21/2021, Machine Learning For Kids had problems to return the confidence percentage, always showing 100% confidence in all tests performed (correct and wrong). Therefore, to verify the real confidence of this judge, the general average was used, through the number of hits and errors collected for each vowel. So, the following data were collected:

Results Analysis - 08/21/2021				
Gesture	Tests Done	Corrects	Wrongs	Hit Average
A	20	12	8	X%
E	20	15	5	X%
I	20	18	2	X%
O	20	12	8	X%
U	20	17	3	X%

Table 1: Results Analysis

It was noticed that the highest error rate was with vowels A and O, a possible justification is that, graphically, the location of the points is similar. In addition, the errors in the other vowels occurred mainly when the hand was at a distance greater than 1 meter from the webcam. Therefore, 50 more gestures were added to the training database (totaling 100 cases of examples per vowel), and new results were obtained:

New Results Analysis - 08/22/2021				
Gesture	Tests Done	Corrects	Wrongs	Hit Average
A	30	28	2	X%
E	30	29	1	X%
I	30	30	0	X%
O	30	29	1	X%
U	30	30	0	X%

Table 2: New Results Analysis

With these more accurate results and with a percentage of correct answers above 90%, it is concluded that this application has high reliability in judging vowels by the Brazilian Sign Language made by hand.

5   Conclusion

This application proved to be effective in judging whether gestures made with one hand obey a vowel pattern from the Brazilian Sign Language with an average precision of X% per gesture.

6   References

...