Hand Gesture Recognition

DATA 606: Capstone Project in Data Science

Daksh Intwala
Jaydeep Radadia
Kratika Aggarwal

Problems

The hearing impaired would benefit from greater accomodations aid in the difficulties they face.

01

Lack of opportunities

Fewer educational and job opportunities due to impaired communication.

02

Social Withdrawal

Reduced access to services and communcation barrier with peers.

03

Emotional Issues

Low self esteem and confidence.

Who are our target audience?

The hearing, speaking impaired

Expressing oneself

It is extremely expensive & time consuming for everyone to learn the language. This will help in live communication wherein the ASL user will be able to directly communicate with the everyone.

Mass Communication

Speech generation is possible which can give voice to the ASL user. It expands the scope of communication for the conversation.

Scope of Language

ASL user can communicate in multiple languages if the text output is translated in different languages using transfer learning (BERT-Hugging Face).

Literature Review

PROJECT - 01

CNN - Letter detection/recognition (93%)

Source: https://www.kaggle.com/code/danielmarom/sign-language-using-cnn-93-with-new-images

Methodology:

a. Read dataset
b. Divide it into train/test
c. Sequential model - relu +softmax
    i. Loss = sparse_categorical_crossentropy
    ii. Optimizer = adam
    iii. Metric = accuracy

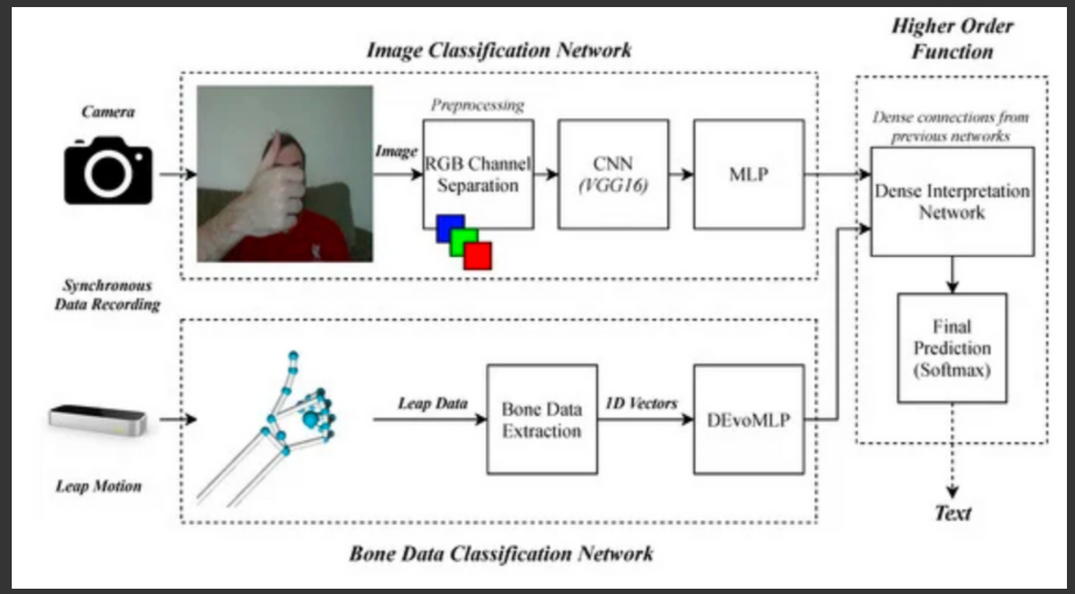
Literature Review

PROJECT - 02

British Sign Language Recognition via Late Fusion of Computer Vision and Leap Motion with Transfer Learning to American Sign Language

Source: https://www.mdpi.com/1424-8220/20/18/5151

Methodology:



Literature Review

PAPER - 01

Hand gestures for emergency situations: A video dataset based on words from Indian sign language

Source: https://www.sciencedirect.com/science/article/pii/S235234092030910O

Methodology:

In this paper, the author has analyzed the images of hand gestures indicating sign language using GoogLeNet CNN model coupled with LSTM architecture and evaluated the performance using F, Recall and Precision scores.

Literature Review

PAPER - 02

Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison

Source: https://openaccess.thecvf.com/content_WACV_2020/papers/Li_Word_level_Deep_Sign_Language_Recognition_from_Video_A_New_Large-scale_WACV_2020_paper.pdf

Methodology:

Here, two different deep learning approaches are used to detect sign language from a video i.e.,

(i) holistic visual appearance based approach, and

(ii) 2D human pose based approach.

Data Source

WLASL

It is the largest video dataset for Word-Level American Sign Language (ASL) recognition

It was created with the intention of facilitating the research in sign language understanding and eventually benefit the communication between deaf and hearing communities.

LINK

https://www.kaggle.com/code/risangbaskoro/reorganize-video-data-wlasl/data

Data

Word : Accomplish

5.4 GB

DATA VOLUME

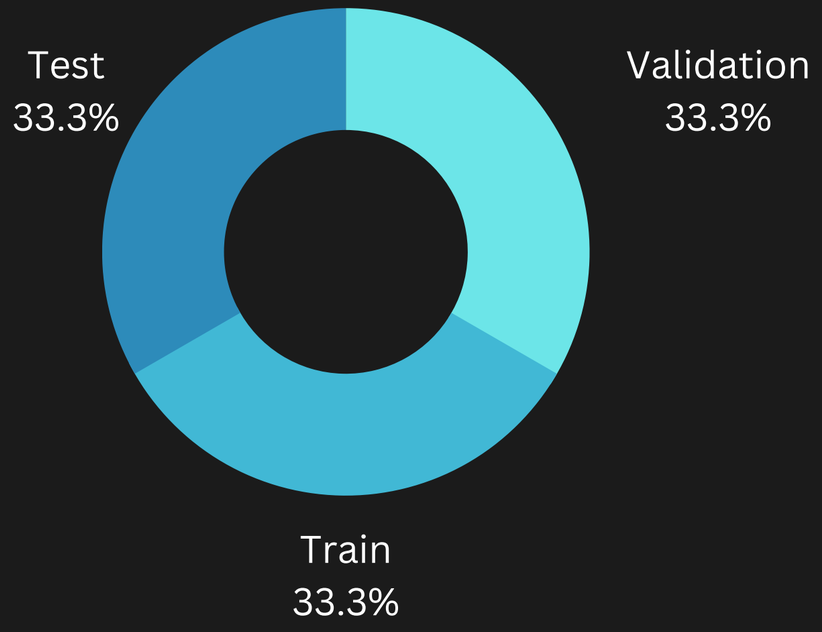
~5

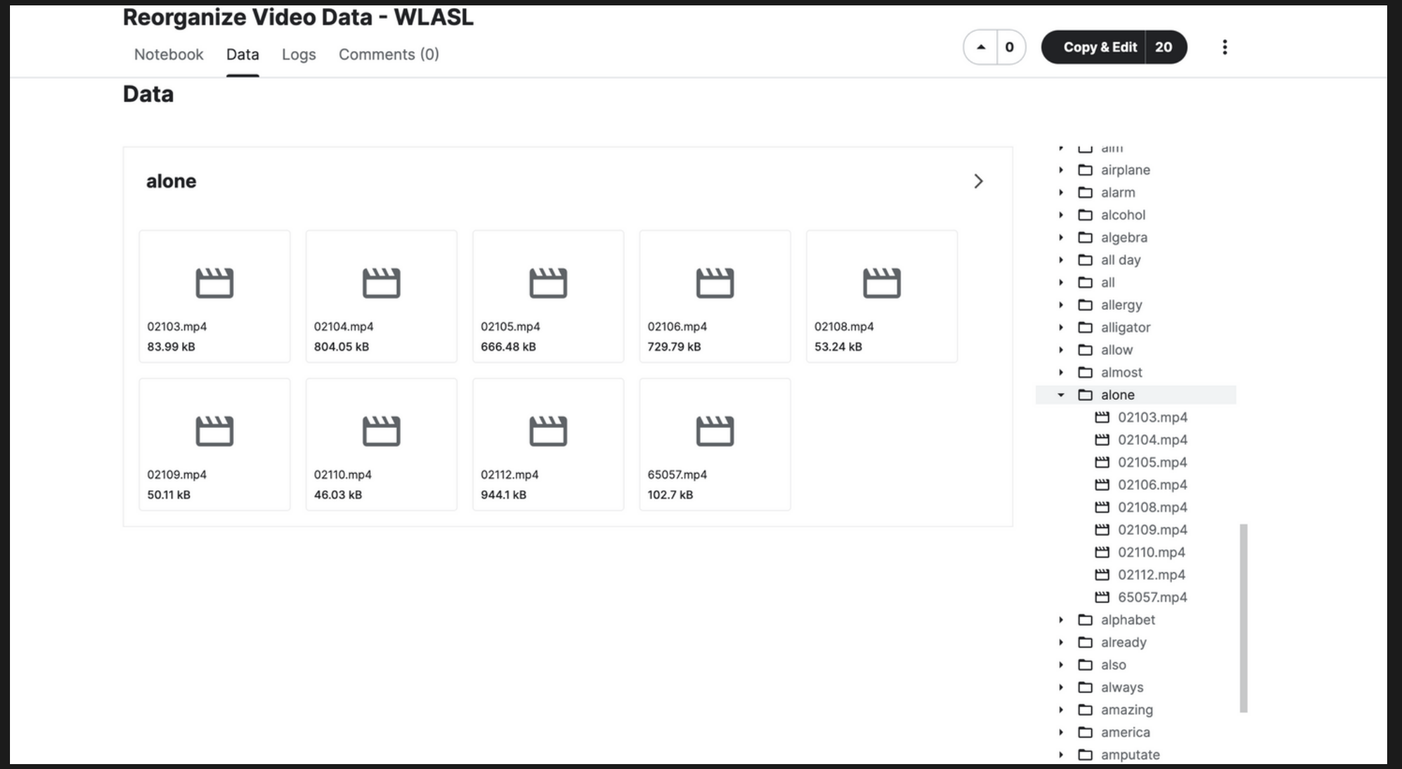
CLIPS PER WORD

2000

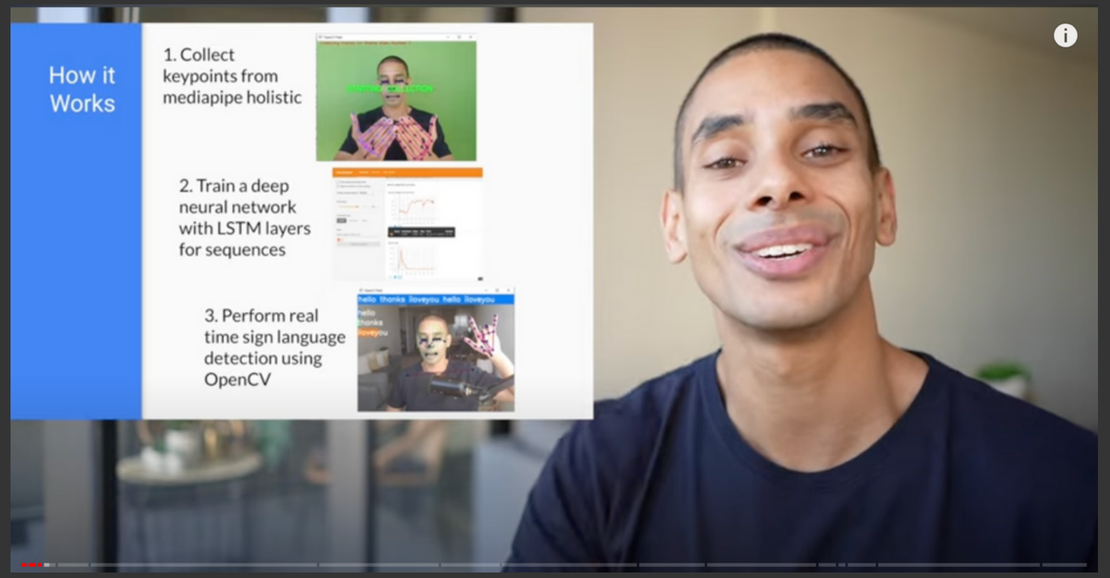
WORDS

Data Preparation





Methodology - Part A



Algorithm - Classification

STEP 1
Using Mediapipe collect keypoints

STEP 2
Train LSTM Model

STEP 3
Perform real-time sign language detection using OpenCV

STEP 4
Model Evaluation: RSME, F1, Recall, Accuracy, Precision

STEP 5
K-Fold Cross Validation (50 Fold)

Methodology - Part B

Processing & Text Generation

STEP 6

RNN Action Predictions

STEP 7

Queuing actions for 2-3 sec frame

STEP 8

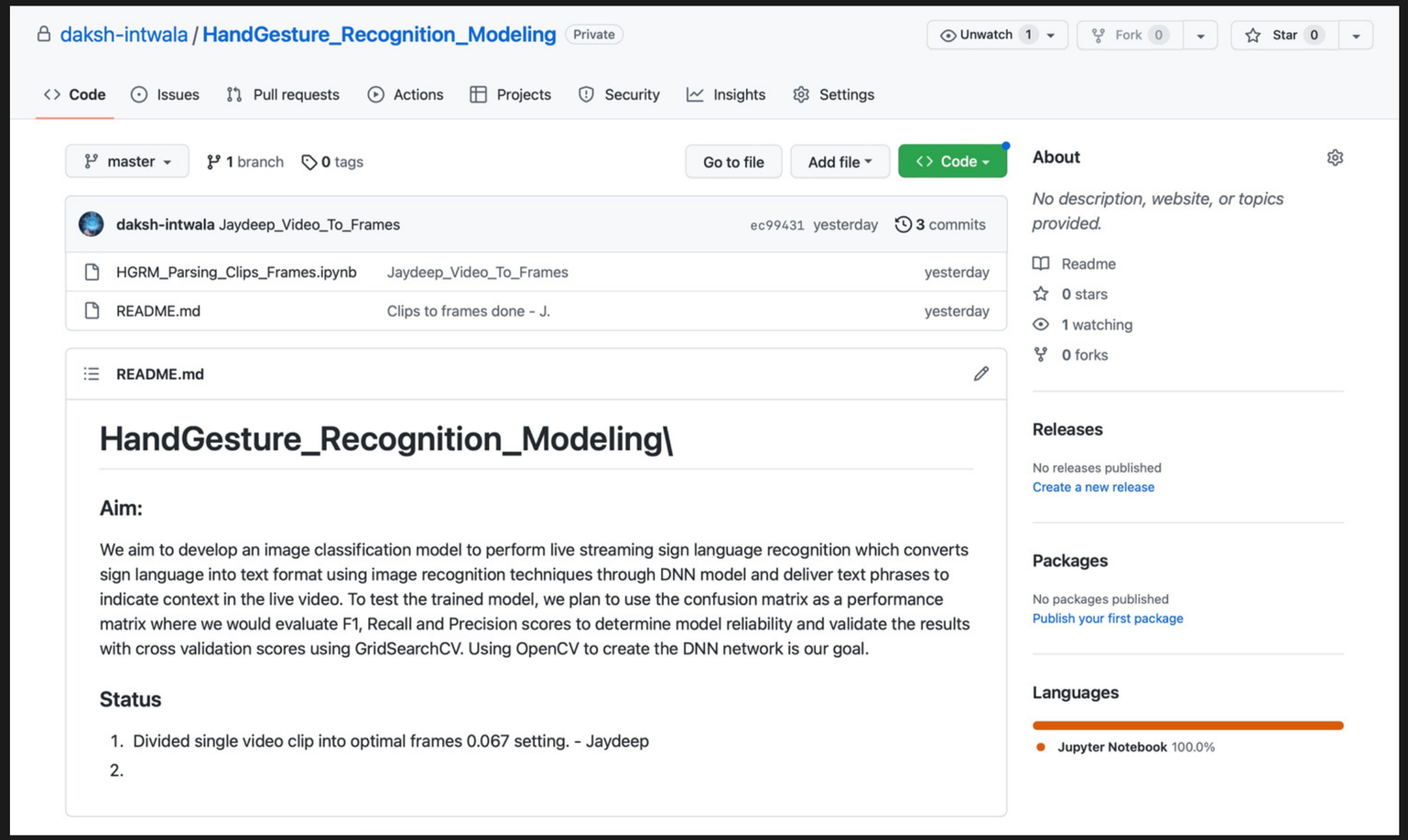
Generate summary through NLG (NLTK, SpaCy)

STEP 9

Grammar-Check Gingerit API

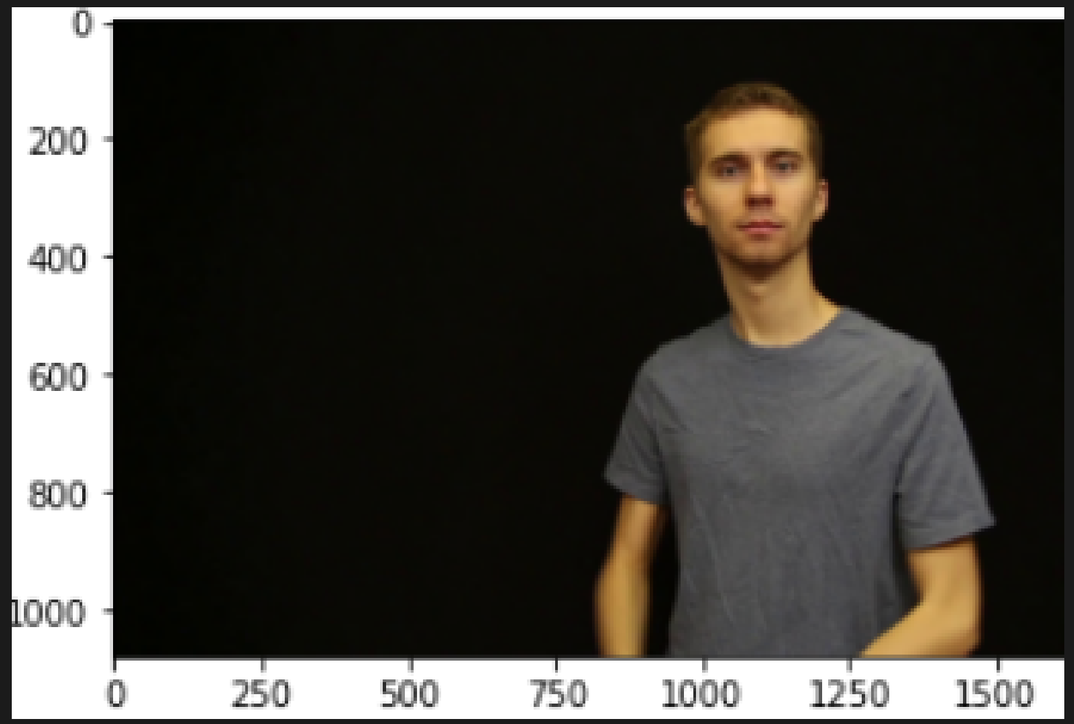
GitHub

https://github.com/daksh-intwala/HandGesture_Recognition_Modeling



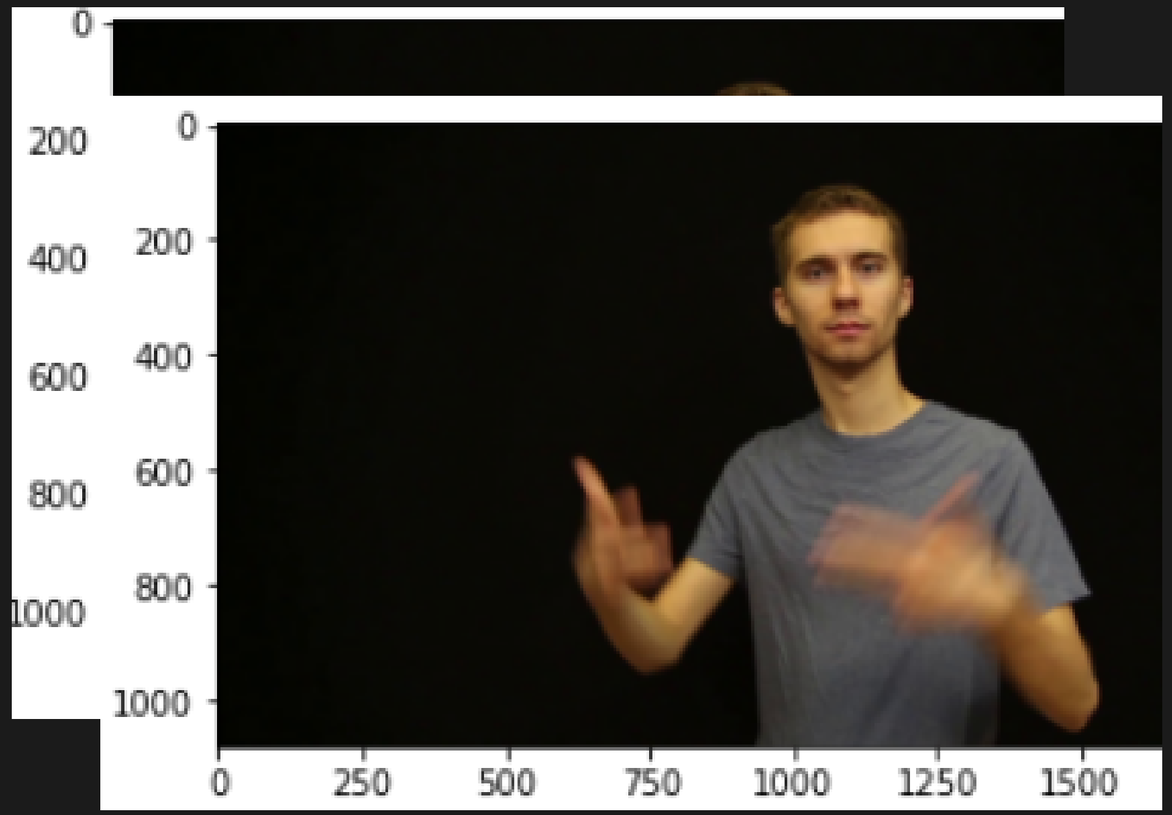
Pseudo Code

1. Parse the videoclip using openCV
2. Determine the optimum frame rate
3. Capture the frames using the .set() method
4. Save the frames



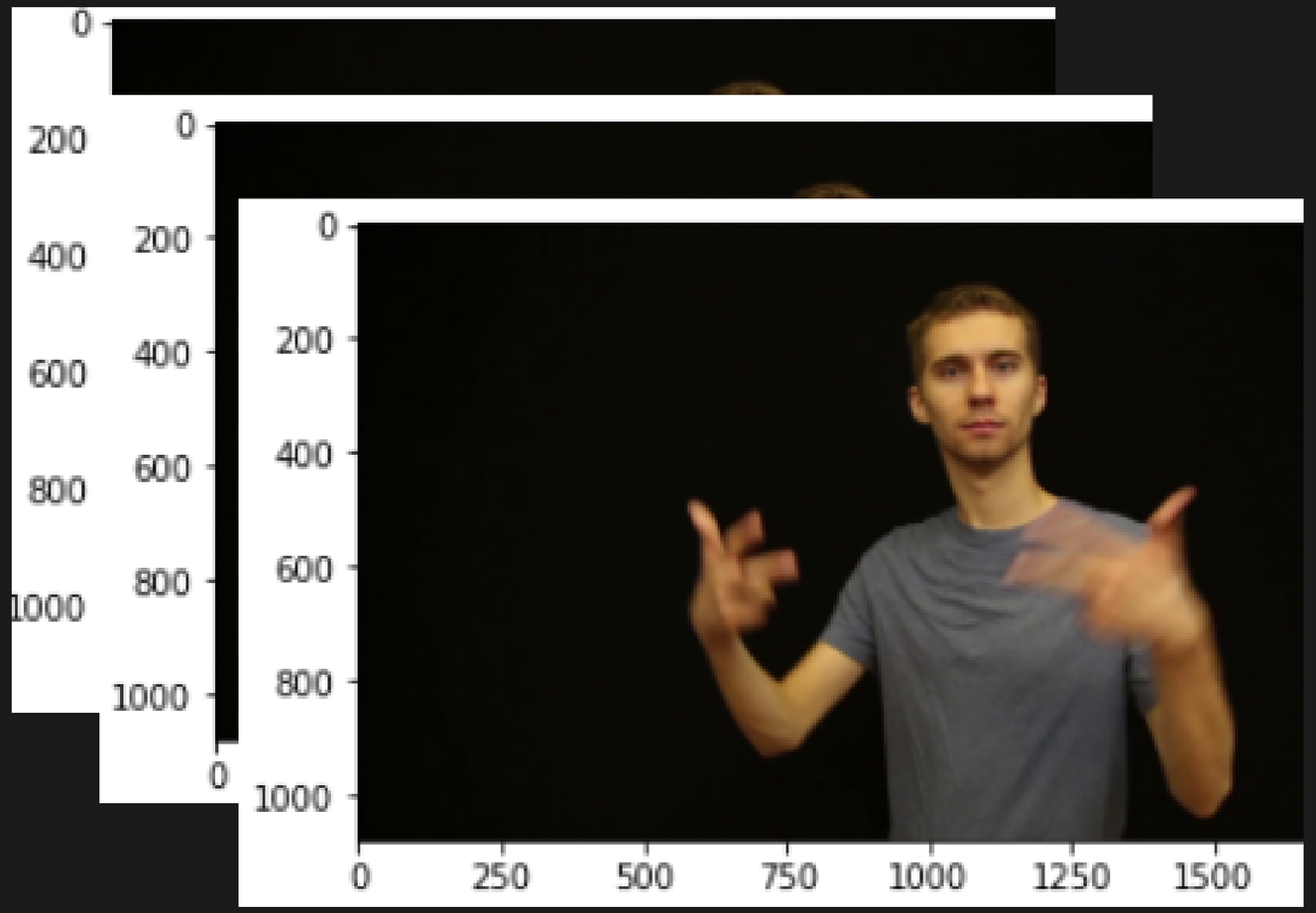
Pseudo Code

1. Parse the videoclip using openCV
2. Determine the optimum frame rate
3. Capture the frames using the .set() method
4. Save the frames



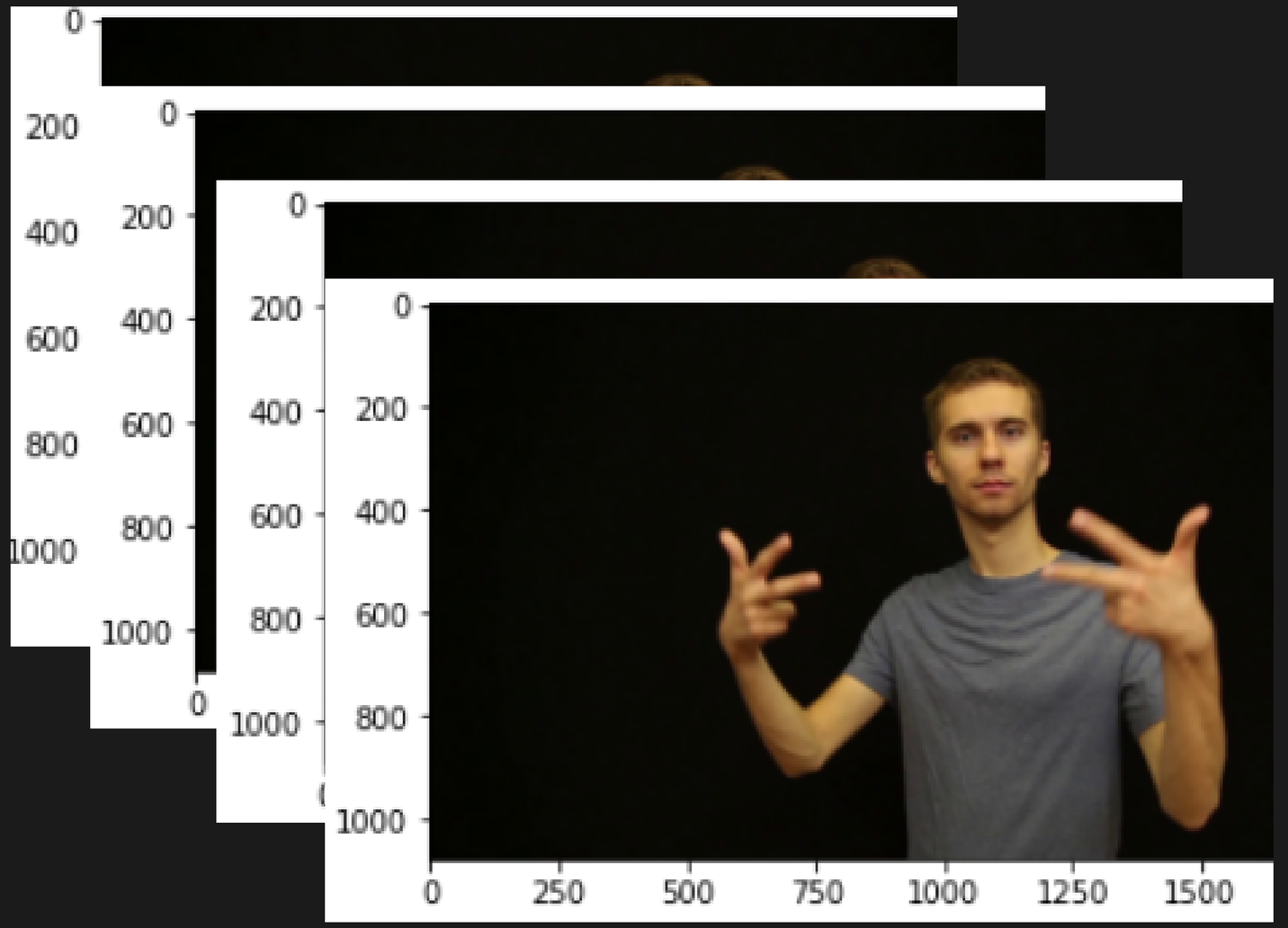
Pseudo Code

1. Parse the videoclip using openCV
2. Determine the optimum frame rate
3. Capture the frames using the .set() method
4. Save the frames



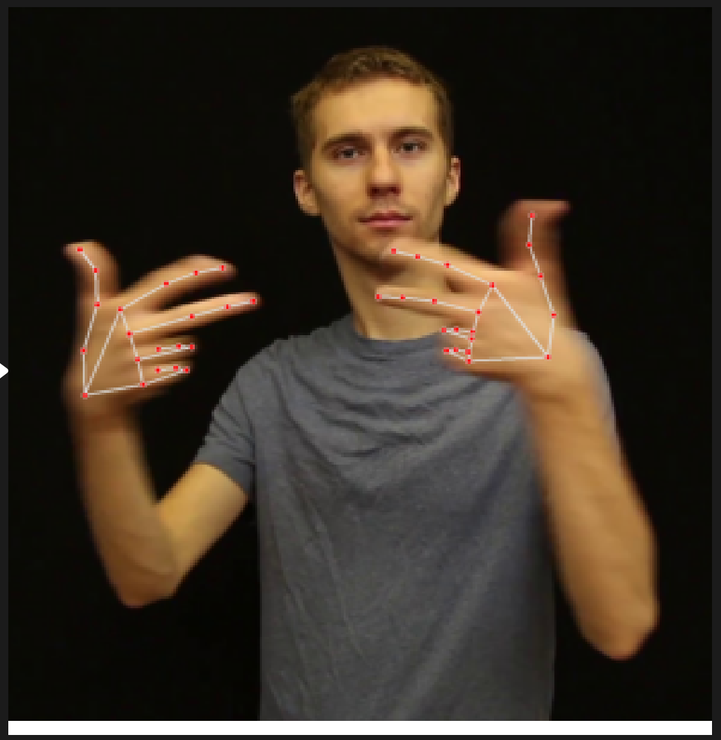
Pseudo Code

1. Parse the videoclip using openCV
2. Determine the optimum frame rate
3. Capture the frames using the .set() method
4. Save the frames



Landmarks

```
Hand No. 1
WRIST:
x: 667.0050430297852
y: 612.0800757408142
z: 0.00073189872637158 26
THUMB_CMC:
x: 665.6389045715332
y: 548.7694072723389
z: -9.656334221363068
Hand No. 2
WRIST:
x: 1333.542366027832
y: 557.2280859947205
z: 0.00044899124077346 64
THUMB_CMC:
x: 1341.1071395874023
y: 497.4174106121063
z: -29.865333437919617
```



References

1. https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/hearing-loss-how-it-affects-people