Software Requirements Specification for Mechatronics: ASL Translator

Team #20, Team Name
Robert Zhu zhul49
Zifan Meng mengz17
Jiahui Chen chenj194
Kelvin Huynh huynhk12
Runze Zhu zhur25
Mirza Nafi Hasan hasanm21

October 4, 2022

Contents

Revision History

Date	Version	Notes
Date 1	1.0	Notes
Date 2	1.1	Notes

1 Reference Material

1.1 Terms, Abbreviations, and Acronyms

Term, Abbreviation, or Acronym	**Description**
A	Shorthand for Assumption
ASL	Shorthand for American Sign Language. It is a form of sign language primarily used in the US and in parts of Canada
CFR	Shorthand for Camera Functional Requirement
CMC	Shorthand for carpometacarpal. This is the joint that connects your thumb to the rest of your hand
DIP	Shorthand for distal interphalangeal. This is the joint on your finger just before where your fingernail is
IP	Shorthand for interphalangeal. This is the joint just before where your fingernail on the thumb is situated
MCP	Shorthand for metacarpophalangeal. This is the joint situated roughly where your knuckles are
ML	Shorthand for Machine Learning
MLFR	Shorthand for Machine Learning Functional Requirement
NFR	Shorthand for Non-Functional Requirement
OpenCV (shortened to CV)	A library of programming tools that enable real-time computer vision. OpenCV provides a basic infrastructure for computer applications that require the use of cameras
PIP	Shorthand for proximal interphalangeal. This is the next joint up your finger from where the knuckles are
TensorFlow	An open-source framework developed by Google, which enables machine learning, deep learning, and other statistical and predictive analytics

2 Introduction

2.1 Purpose of the Project

The purpose of our project is to create a device that will translate sign language gestures into their corresponding words or phrases. This will require the creation and development of a computer vision system alongside a machine learning model that will be used to recognize the hand motions, as well as a Raspberry Pi that will speak the word or phrase. The user will perform the sign language motion that will be captured by our computer vision system through a camera, and processed by our machine learning model and spoken through our Raspberry Pi.

2.2 Scope

2.2.1 In-Scope

The goals for our project are listed in the following table. The primary goals for our project include

- Accurate hand motion recognition: Tracking and recognition of the user's hands
- Real-time translation: Recognition and translation of user's hand gestures with minimal delay

Goals	Desciption
Reliable and Accurate Translations	The Sign Language Translator requires extensive training on the sensors to capture precise hand motion and ignore any human error on the user's part. The processing unit should be able to identify each letter within the American Sign Language using the data collected and transmit dialogue accurately to the user's request.
Real Time Translations	User's should never be required to wait an extensive period of time for the device to process their hand motion and provide a translation. The Sign Language Translator should simulate a real time conversation between regular people to deliver a seamless transition for other parties during presentations or social interactions.
Ease of Use	The user experience is crucial for a communication device. The Sign Language Translator should require minimal time and effort to set up. Once set up, the device should not require much maintenance or updates. Most importantly, the device should not hinder the user's ability to perform the gestures and hand motions of sign language.
Affordability	The Sign Language Translator should be affordable for the end users as to reduce the need of requiring an actual translator to accompany the user during their tasks. The device should remain functional whenever it is required to be used, and the hardware components of the device should be simple and cost-effective.
Customizable to User	As with language, different people might have a certain way of pronouncing a phrase or word and likewise the same could be said with Sign Language with slightly different gestures. The device should be able to adapt to the user and recognize the unique motions instead of forcing the user to slow down for the device.

2.2.2 Out-of-Scope

The stretch goals for our project are listed in the following table. These goals are out-of-scope for our project. They may or may not be achieved depending on our progress and time remaining in the academic year.

Stretch Goals	**Desciption**
Portable	The final device, while requiring OpenCV to scan and process hand motion, should become more portable and lightweight for the user to move around, so as to not interfere with the user's regular activities. The translator text to speech should become an application on all phone brands as for any user with the required equipment to be able to begin using.
Expanding to Different Languages	As a universal sign language does not exist at the moment, there exists deaf/mute individuals who use another form of sign language other than the American Sign language. These include the British, Australian and New Zealand Sign Language (BANZSL), the Chinese Sign Language (CSL), Arabic Sign language, and much more. The device should be able to understand and translate these new hand motions and generate a translation in their native language for this product to be used on a global scale.
Sign Language Education	The final device should be able to recognize the different hand motions and gestures of sign language in order to accurately translate them. This would make the device an excellent educational tool for those looking to learn sign language. The device could provide feedback and tell users how to improve their gestures using it's accurate hand tracking to help teach those unfamiliar with sign language.
Non-real Time Translations	The final product should be able to extract and recognize hand gestures from photos or videos uploaded by the users. In this case, if the users find online photos or videos related to sign language, they can upload them to application/software to acquire text-based translation. This could help the users learn sign language from online sources.

2.3 Usual Operations

The ASL translator will translate sign language gestures into their corresponding words or phrases to help the daily communication for people who have hearing problems. The camera on the device will actively detect the location and the motion of the hand gestures of the users standing in front of the camera. Then the computer vision and machine learning algorithm in the device will accurately translate the motion of the hands and output the correct English words or phrases. Then, the speaker on the Raspberry Pi which is connected to the translator will "speak out" the words or phrases.

2.4 Users and Stakeholders

3 Project Constraints

3.1 Constraints

The project is constrained by the following:

- Translate a subset of the American Sign Language (ASL)
  - Training a machine learning model to encompass the entire ASL would be very time-inefficient for the time constraint of the project
- The project expenses cannot exceed $750 CAD
  - Additionally the project cannot be an off-the-shelf solution and be cost-efficient to attain our goal of affordability
- The project must be completed during the course of the academic year
  - This serves as a time constraint for the project and is also a requirement set by the course

3.2 Assumptions

4 Context Diagrams

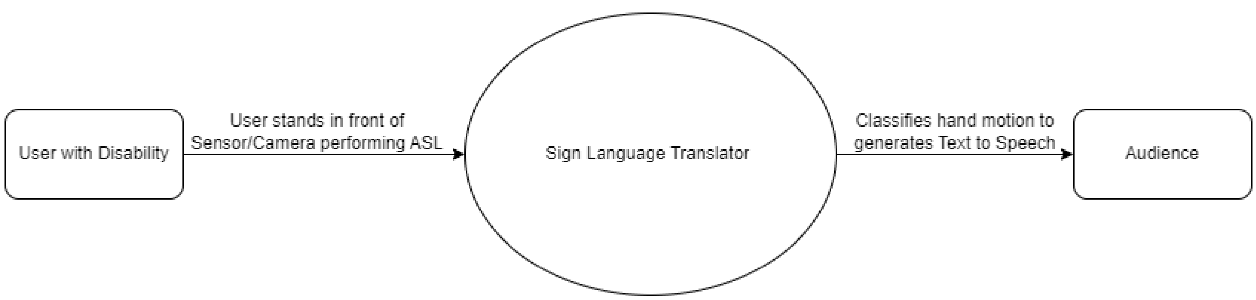

Figure 1: Context Diagram

5 Functional Decomposition

5.1 Data Flow Model

5.2 Monitor and Controlled Variables

6 Functional Requirements

6.1 Camera Functional Requirements

Identifier	Requirement	Rationale
CFR1	User hand gestures should be recognized and converted into input for the system	This is the primary and only way that the end user engages with the system. This is to ensure that their signing is picked up by the camera within a certain degree of accuracy
CFR2	The camera must be able to relay its vision back to the program	This enables validation and testing on the development end. In addition to understanding what needs to be corrected to ensure accurate user input

6.2 Machine Learning Functional Requirements

Identifier	Requirement	Rationale
MLFR1	The program should be able to recognize user hand joints	This enables the ML model to compare and match user hand gestures to ASL
MLFR2	The program should output the x, y, and z coordinates of each joint relative to the camera	This will enable system calibration and aid in enhancing predictive accuracy as the training data set will be primarily static images in contrast to the dynamic input from the end product
MLFR3	The program should recognize up to two hands in the input	The complexity of sign language calls for two hands to enable effective communication. Tracking one hand should also be considered as there are words in sign language that require the use of a single hand
MLFR4	The program should be able to process data in real-time	The translator should relay the relevant translation within a reasonable amount of time to ensure conversation fluidity
MLFR5	The program must be able to calibrate the camera	This is to ensure that the image being processed is undistorted and recognizable to the program to prevent inaccuracies and incorrect output from the ML model
MLFR6	The program should be calibrated to match the speed of the signer	The translator should be able to keep up with the user or the likelihood of a mistranslation will increase
MLFR7	The ML model should be easily trainable	This is how the ML model should learn sign language to use in processing. Making it easily trainable should enable expandability as well. In addition, this will enable the program to adapt to users' specific signing habits and allow for manual correction for the future

7 Functional Requirement Change Likelihood

7.1 Camera Functional Requirements

Identifier	Likelihood of Change	Rationale	What May Be Changed
CFR1	Unlikely	Input component of the system	Input may be changed to sensor instead of a camera
CFR2	Very unlikely	Enables testing and validation for the system	N/A

7.2 Machine Learning Functional Requirements

Identifier	Likelihood of Change	Rationale	What May Be Changed
MLFR1	Very unlikely	Key processing component of the system	N/A
MLFR2	Very unlikely	Enables testing and validation for the system	N/A
MLFR3	Unlikely	Subject to time constraint. ML model accuracy may be sub-par for two hand input	Tracking might only be possible with one hand
MLFR4	Unlikely	Subject to time constraint. Refer to 2.2.1	Data processing in real-time may be difficult and delays might have to be used to ensure translation is as accurate is possible
MLFR5	Very unlikely	Key processing component of the system	N/A
MLFR6	Unlikely	Key implementation aspect. Refer to 2.2.1	Dependent on the processing speed of the program. The speed at which a user can input sign language might be reduced consequently
MLFR7	Likely	Key implementation aspect. But expandability of the program is subject to time constraint and memory required	The ML model may only accommodate a subset of ASL for the sake of time constraint and space saving

8 Non-functional Requirements

8.1 Accuracy Requirement

8.2 Useability Requirement

8.3 Portability Requirement

9 References

10 Appendix

10.1 Reflection