Proceedings of 12th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2022 Month: October Volume: 12

English Sign Gesture Recognition using CNN

Sushant Bhattarai [a], Yateesh Bhurtel [b], Bipin KC [c], Sudarshan Subedi [d]

[a, b, c, d] *Department of Electronics & Computer Engineering, Himalaya College of Engineering, IOE, Tribhuvan University, Nepal*

[a] sushantbhattarai998@gmail.com, [b] yateeshbhurtel@gmail.com,
[c] byapinkc@gmail.com, [d] sudarshansubedi@hcoe.edu.np

Abstract

The understanding of sign language is one of the most exciting fields of gesture recognition study. In this study, camera vision-based methods are employed for image analysis. The goal of the project is to develop sign language system that can identify typical English phonetics as well as a variety of gestures that can be converted to text and subsequently to spoken sentences. A different type of computing technology called convolutional neural networks (CNN) seeks to replicate how the brain operates. An audio built-in system is also used to improve communication between the general public and the community of hearing impaired people. Text-to-speech (TTS) synthesis, a method for automatically converting text into speech, is used and it is utilized in this study to read aloud letters and words that the system has recognized.

Keywords

Gesture Recognition Systems, Convolutional Neural Network, Text to speech conversion, Histogram

1. Introduction

In its various applications, including multimedia computing, secure data transmission, bio-metrics, remote sensing, texture comprehension, pattern recognition, content-based retrieval, compression, and many more, image processing is a rapidly expanding field. This is all about how a computer can detect graphical information following picture processing. Pointing gestures are particularly interesting for communication and may be the most natural interface for choosing among the range of gestures used by humans to communicate with one another. They create the potential for naturally designating locations and items. Given that pointing motions may be used to convey locational parameters in spoken assertions, this is especially helpful when used in conjunction with speech recognition. For disabled people who are incapable of talk and hence unable to communicate, this innovation can be a blessing. It can also be employed as a translator if the speaker and the recipient speak different languages. It has always been difficult to create a user interface that allows for natural interactions between humans and technology, similar to how people normally engage with the outside world. A manual form of communication called sign language was created as a spoken substitute for those with hearing impairments. Wherever there is a need for hearing-impaired communication, sign languages grow. Facial emotion, body language, hand gestures, hand placements, and shapes are all included in sign language. Those who have hearing loss frequently communicate via sign language.

The processing capability of computers have doubled during the last ten years, but the HCI has not much evolved. We are limited by intermediary tools when using a computer (keyboards and mice). These, though, are a hassle and have clogged up human-computer contact. We use voice to interact with one another every day, and we also use gestures to direct, accentuate, and navigate. They are the most comfortable and preferred way for people to communicate with computers. Computer programming and language technology research on gesture recognition seeks to use mathematics algorithms to study human gestures.Gesture recognition might be viewed as a method for computers to start comprehending body language, creating a stronger connection between technology and people. Without the need of any mechanical apparatus, gesture recognition enables human-machine interaction. Computer vision and techniques for image processing can be used to

perform gesture recognition. A mobile phone or computer that senses input can read hand gestures. It interprets how the body moves and communicates with the computer, which uses these motions as input. Then an algorithm that uses statistical analysis or machine intelligence approaches is used to understand these gestures. The creation of a system that can recognize particular human hand movements and use them to communicate information is the main objective of gesture recognition research. It can be useful to communicate with dumb and deaf persons by understanding their hand signals. It facilitates swift action at that precise moment. Many researchers have experimented with various tools and equipment to monitor hand movements, such as gloves, sensors, or wires, but these techniques need the user to wear the gadget, which is useless in real-world applications. Therefore, people considered a method of gesture identification that required no physical contact and would be just as natural as human-to-human connection. Typically, gestures are thought of as hand or body motions that can transmit a person's information to another. Since we are primarily focusing on hand gestures, a gesture is any movement of something like the hand that conveys or accentuates an idea, sentiment, or attitude. According to the many application contexts, hand gestures can be divided into four categories: conversation gestures, control gestures, manipulating gestures, and communicating gestures. Conversational gestures are used in this study to help us communicate more effectively in daily life. A virtual environment can be navigated using controlling gestures. By pointing in the direction of the south, for instance, we can instruct the system to drive a vehicle there. To engage with virtual items naturally, use manipulative gestures. A significant example of a communicative gesture is sign language. It is how hearing impaired individuals communicate with one another. It is suitable as a testing ground for hand recognition systems since it is objective, clearly defined, and rarely gives rise to ambiguity. The two main categories of these methods are "Data-Glove based" and "Vision based." The data-glove is a pricey connected electronic gadget that serves as the foundation for many identification systems. Several sensors are positioned on the glove to track the hand's overall location and relative configurations. The cost of the glove is one restriction of this strategy, and another issue is that it requires a physical connection between the user and the machine, here the camera. Due to this drawback, more and more academics are becoming interested in vision-based technologies, which are remote and just require one or more cameras to operate. In this research, a camera is employed in conjunction with a vision-based system. Digital image processing can eliminate issues like the accumulation of noise and channel distortion during processing and allows a much wider choice of methods to be employed to the input data.

1.1 Problem Statement

People with the appropriate abilities can express their thoughts and opinions verbally. For those who have trouble hearing or listening, sign language is the sole available form of communication. They constantly have issues communicating with persons who have disabilities. They struggle to communicate their concepts and thoughts to regular people who have little to no knowledge of sign language. These people communicate through sign patterns rather than acoustic sound signals. These indication patterns are created by combining the shapes of the hands, the direction and motion of the arms, the torso, and the face, as well as the patterns of the lips. Due to this, the community of persons who have trouble hearing and speaking loses interest in shared activities, sometimes avoids speaking to able-bodied people, and ends up living alone.Their whole growth and quality of life are affected by this. There are numerous sign language recognition systems that have been created to address this issue, but more precise and efficient systems are still required. There are distinct sign languages for each nation and each location. The earlier researchers' solutions currently rely on changing an action-based verb into an equivalent symbol. These systems are only capable of processing a certain amount of action verbs in a given language. The goal of this research is to create an English alphabet sign language recognition system. People with speech and hearing impairments should find it simpler and more user-friendly to use the suggested system. The suggested system will translate sign language into its text equivalent.

1.2 Aims and Objectives

To recognize sign languages using CNN and translate the subsequent sign language into letter and then to word

1.3 Scope and Application

This platform can easily be operated by people who are familiar with sign languages. There are sectors where this Research can be used and by people with special disabilities and is most effective for them. Some of the applications are:

i. Education sector
ii. Office reception navigation
iii. Chatting applications in messenger

2. Related Works

The study on key components for creating a system to recognize sign language is reviewed in this section. The identification of hand sign language has been the subject of extensive research in recent years. Instrumented gloves and vision-based technology are both employed for sign language recognition. The study concentrates on a method of hand movement detection that leverages vision.

i. Vision-Based strategies Although vision-based methods are straightforward, many gesture challenges arise due to complex backgrounds, varying lighting, and other objects with different skin tones in close proximity to the hand object. Vision-based methods use only cameras to record the image needed again for natural interaction between humans.

ii. Instrumented Glove approaches: Marked gloves, also known as colored markers, are gloves that the human hand wears and that are colored to guide the process of monitoring the hand, finding the palm, and extracting the geometric elements required to make the hand shape.

Gesture Recognition Techniques: Most studies that study the process of gesture recognition employ ANN as a classifier.

Histogram Based Feature: This is a technique for identifying motions based on orientation histogram pattern detection.

Fuzzy Clustering Algorithm: The key distinction among fuzzy clustering and other clustering algorithms is how sample data are divided into groups in a fuzzy manner in fuzzy clustering. In other clustering algorithms, a single data pattern may belong to multiple data groups.[1]

Hidden Markov Model (HMM): A Markov chain is a mathematical representation that provides information on the probability of collections of random variables, or phases, each of which may take values from a finite set. Words, tags, or characters representing anything, such as the weather, can be part of these collections. A Markov chain makes the extremely strong assumption that the only thing that is important is the current state if we wish to anticipate the future in the sequence. The states that came before the present state have no bearing on the present state's future. It's as if you could only look at the weather today in order to anticipate the weather for tomorrow, but not yesterday.[2]

A highly precise image processing technique called the Artificial Neural Network based Sign Gesture Recognition System was developed to recognize English Sign Language. Gloves or markers are not required for the algorithm's implementation. The system in place makes use of the boundary tracing idea, which also incorporates finger-tip detection, to count the number of opened fingers. The author makes an effort to identify the character produced by the screen sensor and convert it to speech using a recognition algorithm. The author proposes a method for identifying hand gestures based on finger border tracing and tip of the finger detection. [3]

A system that uses MAT LAB to recognize 26 hand movements in Hindi sign language was shown by four people. Segmentation is possible through image processing. The Eigen values and Eigen vectors, which are employed in recognition, are among the features that are extracted. The gesture was detected using the Linear Discriminant Analysis (LDA) technique, and the recognized gesture was then turned into text. [4] Another study on real-time hand-gesture recognition using HMM is given (Hidden Markov Model). Based on the idea that skin tone in photographs occupies a connected volume in HSV space, they developed a method for segmenting skin tone in that area. They also created a system that could identify gestures from segmented hand photos using a back-propagation neural network.[5]

The hand regions are extracted from various images obtained by a multi perspective camera system, and the "voxel Model" is then constructed. Two people proposed this innovative methodology as a hand-pose estimation that may be utilized for vision-based human interfaces. [6], in which they categorized objects using an RBF network and used an image furrier descriptor as their main feature. The overall efficiency of their system was 90.9%. In their technique, Claudia Nolker and Helge Ritter introduced a sign language recognition method based

on recognition of finger tips. Based on this, a 3D model of the hand is produced and neural networks are used.

3. Methodology

3.1 System Block Diagram

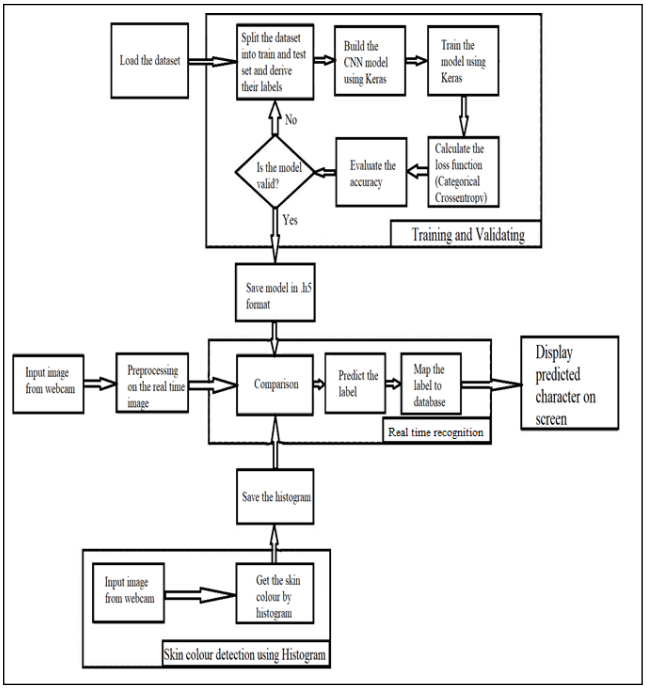

Figure 1: Block diagram of English Sign Gesture Recognition system

3.2 Dataset Collection

The dataset is collected from Kaggle, which was uploaded by a data engineer Dibakar Saha. [7] The dataset was also created by our own and is used for the training of the model. We downloaded 800 images from Kaggle and created a python script that captures 400 images. The Research will use those corresponding datasets with each image with dimensions 75 x 75 pixels. There are 26 classes for the letters A-Z. The total dataset is divided in a 80%-20% manner which consists of 80% training set, and remaining 20 % is divided into 10% testing and 10% validation sets. The figure below shows all gestures taken from each gesture folder.

Figure 2: Sign Language Datasets

3.3 Building CNN model architecture

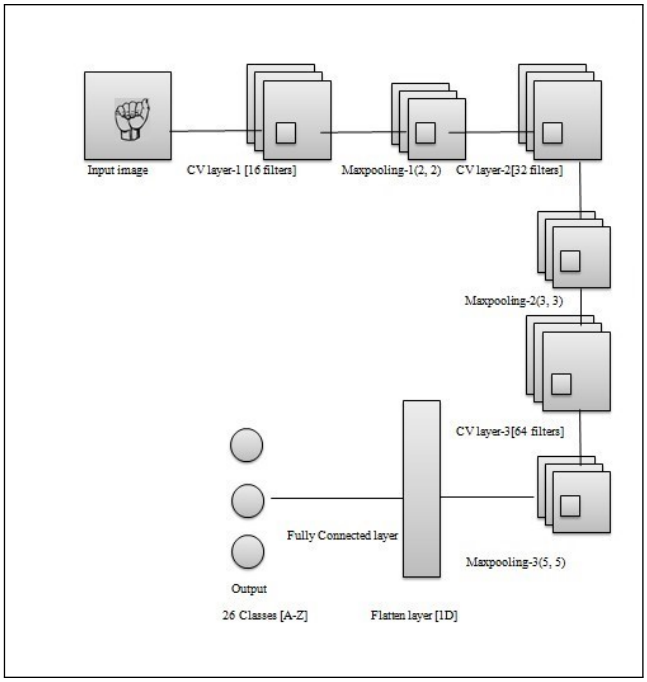

Figure 3: The architecture of Convolutional Neural Network for the system

After the model architecture has been developed, it is trained with batch size of 500 and over 20 epochs. The performance of the model depends on all the above given parameter values. The validation image and validation labels play a very crucial role in evaluating performance. Various graphs is derived while estimating model performance. The epoch vs loss graph, epoch vs accuracy graph and confusion matrix are derived after training the model.

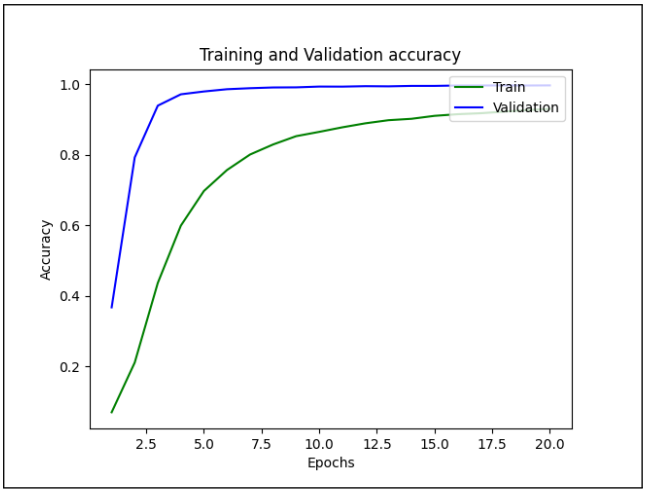

Figure 4: Epoch vs Accuracy graph for 20 epochs

The model's training and validation accuracy versus 20 epochs are shown in the graph. The model's first training accuracy is around 7%, and its initial validation accuracy is around 38%. The training accuracy increases above 80% after epoch 6. Similar to this, during the third epoch, the validation accuracy increases to above 90%.

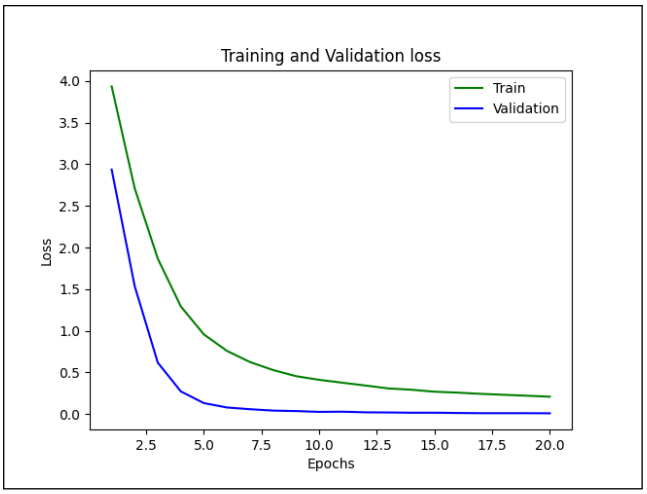

Figure 5: Epoch vs Loss graph for 20 epochs

The graph displays the model's training and validation loss over 20 epochs. The model's initial training loss is close to 3.9, and its initial validation loss is close to 2.9. The training loss decreases to less than 0.5 after epoch 9. Similar to this, after the fourth epoch, the validation loss drops below 0.5.

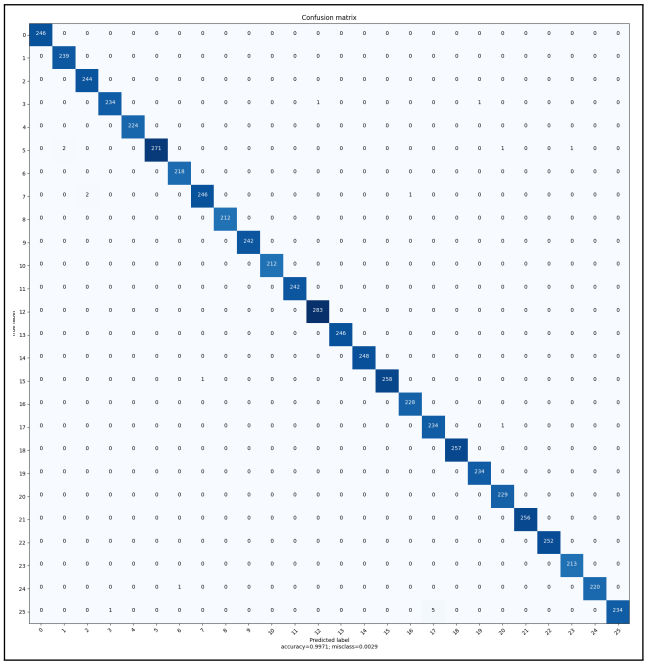

Figure 6: The confusion matrix for 26 classes (from A to Z)

3.4 System Implementation

3.4.1 Setting up hand histogram

Setting up hand histogram is detecting the skin color of the user hand. Histogram must always be set when the lighting condition varies. The following steps are carried out during the hand histogram setup:

Image data preprocessing: This makes the images in standardized format where we have reshaped the image to 75*75.

Building squares for color detection: Fifty squares are built and the hand is kept in the squares in such a way that all squares are covered up by hand. On pressing the button "C", the skin color is detected through those squares. The function "getStructuringElement()" is used that takes the parameter "MORPH_ELLIPSE". This function now shows all the objects in picture as white, if the object color matches the skin color inside of those squares. All the other objects are then shown as black.

Saving the histogram file: The perfect threshold picture is derived and finely tuned by the button "C". Then, after the user gets satisfied, histogram can be saved as a file using pickle library by pressing button "S".

3.4.2 Realtime Image Data Preprocessing

Frames Standardization: The frames must initially be captured through webcam. These frames must then be flipped using "flip" function. Then, the frames are resized using "reshape" function.

Noise removal Noises are removed using Gaussian blur and Median blur functions.

Comparison with histogram: The histogram that was set up and saved as a histogram file is now passed as a parameter to the "calcBackResearch" function. This function only shows the hand as white and background as black, if the histogram was set up properly.

Binarizing / Filtering / Threshold: Binarizing is a process which converts a gray level images to a binary image. Gray level images have 0 to 255 levels. Where in binary images there are only two values: 0 and 1 (black and white). The images are converted to binary images using the "thresh" function.

3.4.3 Comparison with model and acquiring probability

The saved model (of .h5 format) was compared to the real time binarized image using the "predict" functio. The output of this function was used to provide the prediction probability and prediction class. Only the prediction class with prediction probability greater than 90 was used for further processing.

3.4.4 Mapping with the database

The predicted class was mapped with the database (gesture.db), manually created earlier that enlists a table with two columns namely gesture_id and gesture_name, to give out actual character to be displayed onto the screen.

3.4.5 Displaying the output

The output was then displayed on the blackboard in the console. This is done simply through "imshow" function of OpenCV library.

4. Result analysis and Discussion

4.1 Performance of the model

4.1.1 Accuracy and Loss

Parameter	Value
Training accuracy	0.9306
Validation accuracy	0.9974
Training loss	0.2094
Validation loss	0.1040

4.1.2 Realtime Outputs

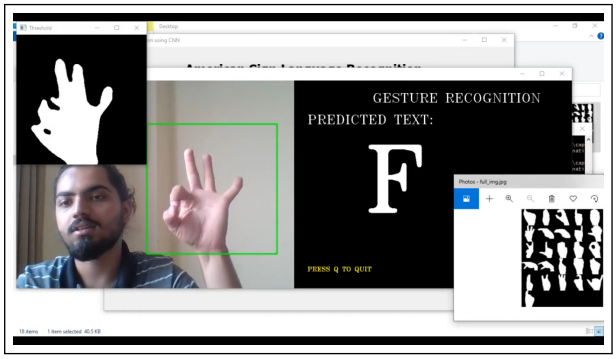

Figure 7: Recognition of letter "F"

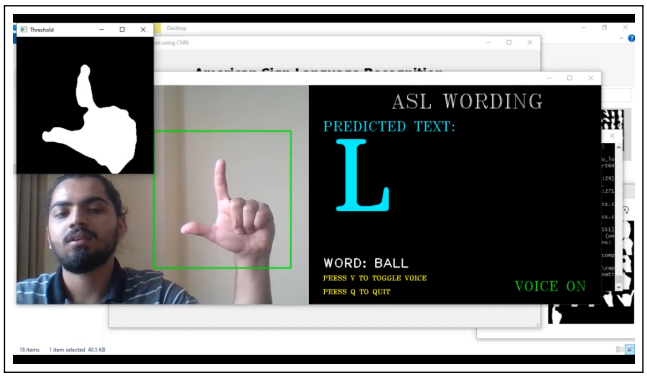

Figure 8: Recognition of word "BALL"

5. Conclusion and Future Enhancements

5.1 Conclusion

People who are hard of hearing will find this research helpful. The user will receive an efficient and accurate display of alphabet letters on the screen, thanks to this research. Based on the study's findings, this research offers a variety of services to users with various demands and lessens the inconvenience that comes with communication. The goal of this research was to increase the likelihood that letter from A to Z movements would be recognized. For instance, the letter "C" is quickly recognized to demonstrate how

this work not only increases the rate at which some alphabets are recognized but also preserves the rate at which others are recognized. It was noted that the recognition rate had increased and that the recognition time had greatly decreased. CNN has been employed to accomplish this.

To sum up, this research approximated the implementation of all the criteria. In order for the system to run well, all of its components typically function correctly. Every test case is designed for every element in every iteration of the implementation, and all test results are double-checked to ensure that everything is functioning as intended.

The research led to the creation of a straightforward desktop application enabling English Sign Gesture Recognition for differently-abled hearing impaired people.

5.2 Limitations

- Only feasible for people who have the knowledge of English Sign Gestures
- Can face worst accuracy in the case of bad lighting conditions
- Bad accuracy in case the background color and hand color are identical
- Bare hands can create noisy gestures. Gloves can be used for fine edges in gestures

5.3 Future Enhancements

This Research is an initial step in reaching the effective solution for the daily concern. This Research can be extended in multiple ways in the future such as:

- Web Application: Users will be able to communicate with each other through camera via web browsers
- Android Application: Users will also be able to communicate through mobile app using messenger
- Sentence generation: The Research can be further converted into recognizing daily communicating sentences

Acknowledgments

This work was supported by Himalaya College of Engineering. The authors have their kind regards to Er. Ashok Gharti Magar, the head of department for Electronics and Computer engineering at HCOE for his priceless guidance and support. The authors are also grateful towards DHOD, Er. Devendra Kathayat, Lecturer, Er. Anil Pudasaini and other faculty members for their precious help and support in the research project ”English Sign Gesture Recognition using CNN”.

References

[1] A. Sethi, S. Hemanth, K. Kumar, N. Bhaskara, and K.R. Rao. Signproan application suite for deaf and dumb. *IJCSET*, 2(5):1203–1206, 2012.

[2] V Nayakwadi and N.B. Pokale. Natural hand gestures recognition system for intelligent hci. *International Journal of Computer Applications Technology and Research*, 2013.

[3] H. Parul. Neural network based static sign gesture recognition system. *International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)*, 2(2):3066–3072, 2014.

[4] M. Srinithi M. Suriya, N. Sathyapriya and V. Yesodha. Survey on real time sign language recognition system: An lda approach. *ICEIET–2016*, pages 2348—-2387, 2016.

[5] H.-K. Lee and J. H. Kim. Interactive learning of gestures for human robot interfaces. 1996.

[6] M. Imai E. Ueda, Y. Matsumoto and T. Ogasawara. Hand pose estimation for vison based human interface. *IEEE Transactions on Industrial Electronics*, 50(4):676–684, 2003.

[7] EvilPort2. Sign language gestures, Oct 2017.