**1.0 Introduction**

**1.1 Background**

Development of language as a communication medium was a huge achievement in evolution, and there is no human community without it. Humans have a natural tendency for language in two different modalities: vocal-auditory and manual-visual (Sandler,2003). Since the beginning of civilizations, vocal-auditory modality was the predominant method. According to Meadow (2000) speech is the predominant medium for its transmission and it seems that spoken languages themselves are either also very old or are descended from other languages with a long history. On the other hand [Fischer](https://www.researchgate.net/profile/Susan_Fischer4) (2015) stated that sign languages do not have the same histories as spoken languages because special conditions are required for them to arise and persevere, and for this reason they can offer unique insight into essential features of human language. The [visual](https://en.wikipedia.org/wiki/Visual_language) recorded history of sign language in Western societies starts in the 17th century, as a [language](https://en.wikipedia.org/wiki/Visual_language) or method of communication (Sandler, Wendy, Martin (2006). At present, sign language can be interpreted as a unique system of conventional gestures, mimic, hand signs and finger spelling, plus the use of hand positions to represent the letters of the alphabet, ideas or phrases (Bauman & Dirksen,2008).

Sign language may be categorized into two types. “The first is used by individuals who have accesses to vocal-auditory language and the signs are used for special situations, such as in military communication and when practicing monastery rituals.

The second is used by those who do not have access to vocal-auditory language, namely the deaf(Ruben,2005) and deaf-mute people. Deaf-mute is a term which was used historically to identify a person who was either deaf using a sign language or both deaf and could not speak.

Many natural languages have created their own sign language system with different grammar, syntax, and vocabulary. Each displays the kinds of structural differences from the country’s spoken language that show it to be a language in its own right. (Sacks, 1989)

For example, ASL and British Sign Language are different, mutually unintelligible languages since the American and British Deaf communities were not in contact with each other, the two languages developed independently Perlmutter,2018). French Sign Language, Danish Sign Language, Taiwan Sign Language, Australian Sign Language, Thai Sign Language, Finnish Sign Language, Brazilian Sign Language, and many others have developed in communities of Deaf people, just as spoken languages have developed in communities of hearing people (National Institute on Deafness and Other Communication Disorders,2019)

Among those Sinhala Sign Language is a [visual language](https://en.wikipedia.org/wiki/Sign_language) used by [deaf](https://en.wikipedia.org/wiki/Deaf) people in Sri Lanka. Sri Lankan Sign Language was fully built on the foundation of British Sign Language but have lots of variations (Wikipedia, 2019). Sri Lankan Sign Language currently consists of more than 2000 sign based words.

Presentation of sign language consist of two techniques. They are established or productive signs and fingerspelling. In any sign language there are signs allocated for particular nouns, verbs and phrases. These signs are frequently used and highly standardized. These are known as established signs.

These signs are “frozen” and form the basis of the vocabulary listed in dictionaries of sign language. Productive signs make use of a much larger and more varied selection of locations and movements than established signs. These signs are actively created by signers as they put together combinations of meaningful units. This explains why these are called “productive” signs. These “meaningful units” can be used to extend or modify the meaning of established signs.

Examples of meaningful units are: handshape, hand orientation, sign location and movement, non-manual features, rate, stress, duration and repetition. Productive signs combine different meaningful units in different combinations as the need arises to produce signs that may have never been signed before but can be understood in a particular context. In any given signed conversation there is most probably a significant number of signs which have been created or re-created on the spot as required by the topic or context of the discussion.

Fingerspelling is using your hands to represent the letters of a writing system. In English, this means using 26 different hand configurations to represent the 26 letters of the English alphabet. As such, fingerspelling is not a signed language in and of itself, rather it is a manual code for representing the letters of the English alphabet. Among deaf and mute individuals, finger spelling is more often used in conjunction with sign language for proper names and terms for which there are no signs. It appears that fingerspelling was first used by hearing people to represent the written form of spoken language; however, fingerspelling is now completely integrated into natural signing.

According [Fischer](https://www.researchgate.net/profile/Susan_Fischer4) (2015),fingerspelling is occasionally integrated with established signing, and particularly use when addressing names of peoples and objects.

The only time fingerspelling might be exclusively used is in an educational setting such as the Rochester Method, or with deaf/blind people (or in very dark lighting conditions) whereby the letters are spelled onto the hand of the deaf/blind person. (Jhonston & Schemberi,2007)

According to Sri Lankan Federation of the Deaf, there are over three hundred thousand plus (300,000+) Deaf people in Sri Lanka. Moreover, the World Health Organization has revealed that approximately 9% of the population in Sri Lanka have speaking and hearing impairments. The reasons for these disabilities are not mere due to birth issues. According to Department of Census and Statistics Sri Lanka road accidents, riots and violence and war and terrorism are also among major reasons which creates speaking and hearing impairments.

As a result of that Sign language is an extremely important communication tool for deaf and hard-of-hearing people. There are only few number of people who are competent on Sinhala sign language and hence create a great difficulty for deaf people to engage in their social life and endeavors. Other people have to learn the sign language to communicate with deaf people and same might be useful to themselves due to other unfortunate factors which create speaking and hearing impairments as mentioned above.

There is a lack of interest in the natural persons to learn SSL. And because of that deaf and mute people cannot interact with the normal people and eventually the deaf people get isolated in the society.

It simply takes time to learn sign language when compared to a natural language since it takes communications to a completely different level and demands that you master eye gazing to better navigate the give-and-take of communal interactions. Additionally, there is only few number of people who can teach the Sinhala sign language best.

It is much convenient to both normal persons and deaf and mute people if there is an effective device based real-time translator.

**1.2 Problem Definition**

The existing applications on the topic are standalone learning applications of sign language for a beginner and does not support real time applications. The real time sign language translating applications are still in the research levels for English languages and many other languages. Also for Sinhala sign language. Majority of them are focusing on developing electronic devices. But it requires more power supply and it is very costly. So some researches use the static hand gesture recognition system using digital image processing.

Hand gestures can be different from person to person. The length, size can be different. So have to use machine learning to identify similar hand gestures when using image processing for identify the hand gestures allocated for a particular sign. other applications that are converting Sinhala sign language into text did not use machine language.

The systems that are translating Sinhala sign languages translate words and phrases into text only. There are no systems that are translating Sinhala sign language for Sinhala letters into Sinhala letter texts. Current systems are only capable of identifying few number of words and sentences among 500+ Sinhala sign language. To identify the rest, each word has to be inserted to their system.

Alphabet of Sinhala language only consist of 60 letters. Since letters are the building blocks of any language, converting signs based on letters to text can increase the horizons to identify and translate almost any word.

Therefore, this research is based on creating real time Sinhala sign language translator based on letter based signs using image processing and machine learning with the intention of achieving effective communication platform for people with visual and verbal impairments.

**1.3 Motivation**

Differently abled individuals are common element of any human society. But due to communication barrier they got neglected from the main society. People are reluctant to learn sign language to build up an effective communication platform with deaf or deaf and mute people. To make them feel involved and respected, effective communication bridge has to be adopted between deaf people and rest of the society.

The existing applications on the topic are standalone learning applications of sign language for a beginner and does not support real time applications. The real time sign language translating applications are still in the research levels for English languages and many other languages.

Hence his research is based on creating real time Sinhala sign language translator based on letter based signs using image processing and machine learning. ~~End product may be capable of tracking the hand gestures of Sinhala sign language for letters and print it in a text field on the user’s device.~~ ( can we use “may be” in a research?)

**1.4 Aim and Objectives**

**Aim**

Creating an effective communication platform for deaf and mute people in the society through real time Translation of Sinhala Sign Language into text through recognition of alphabet based signs

**Objectives**

1.Review the prevailing techniques and measures on real time sign language translation

**2 and 3 have to discussed**

4. Develop an application which is embedded with the capability of translating Sinhalese sign language into text through recognition of alphabet based signs.

**1.5 Limitations**

There are 60 letters in Sinhala alphabet. Correspondingly there are 60 signs to denote each letter. There are 2 types of hand signs indications for those letters as static hand signs and dynamic hand signs. This research has been limited only to recognition of static hand signs. (Has to revised)

**1.6 Major achievements**

This research proposes a method to convert letter based hand signs into text using combination of image processing techniques and machine learning techniques.

( At the END of research)

* 1. **Structure of the Dissertation (Thesis? )**

***Chapter one- Introduction***

This chapter gives a general overview of the research and guides to the content of the study while presenting the background of the study, aim and objectives, scope and limitations of the study and methodology adopted.

***Chapter two: Literature Review***

The second chapter includes the previous works for the identification and translation of sign language. It comprehensively discusses about the techniques for image preprocessing.

***Chapter Three: Research Methodology***

The third chapter describes the theoretical concepts of image processing and fuzzy logic behind on the research. Then it explains method of the project which is carried out to perform hard exudates detection successfully. It contains three stages for image preprocessing, exudates detection and hard exudates detection.

***Chapter Four: Analysis and Research Findings***

Findings through interviews have been critically analyzed within this chapter while elaborating how the research objectives were accomplished

***Chapter Five: Conclusion and Recommendations***

The conclusions of the overall research presented in this chapter together with preferred further researches.

**References**

Bauman, Dirksen (2008). *Open your eyes: Deaf studies talking*. University of Minnesota Press. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [0-8166-4619-8](https://en.wikipedia.org/wiki/Special:BookSources/0-8166-4619-8).

[Fischer](https://www.researchgate.net/profile/Susan_Fischer4) S.D, Sign languages in their Historical Context Chapter (PDF Available) · January 2015

DOI: 10.13140/2.1.2085.5683

In book: The Routledge Handbook of Historical Linguistics, Chapter: Sign Languages in their Historical Context, Publisher: Routledge, Editors: Claire Bowern & Bethwyn Evans, pp.442-465

Perlmutter D.M. (2018) *What is sign language ?* The Linguistic Society of America.Retreived from <https://www.linguisticsociety.org/sites/default/files/Sign_Language.pdf>

Sacks, O , , 1989.. Seeing Voices University of California Press, Paperback published by Harper/Collins

Goldin-Meadow, S. (2003). The resilience of language: What gesture creation in deaf children can tell us about how all children learn language. NY: Psychology Press.

Sandler, Wendy and Diane Lillo-Martin. 2006. Sign language and linguistic universals. New York: Cambridge University Press

Australian sign language: An introduction ,2007, Cambridge University press, Jhonston T. Schemberi A.

National Institute on Deafness and Other Communication Disorders.2019 <https://www.nidcd.nih.gov/health/american-sign-language>

**Literature Review**

Sinhala Sign Language is a visual language used by deaf people in Sri Lanka. According to Sri Lankan Federation of the Deaf, there are over three hundred thousand plus (300,000+) Deaf people in Sri Lanka. Moreover, the World Health Organization has revealed that approximately 9% of the population in Sri Lanka have speaking and hearing impairments. So Sign language is an extremely important communication tool for deaf, dumb and hard-of-hearing people. But not every single person in Sri Lanka can recognize the sign language. It simply takes time to learn sign language when compared to a natural language. S. P. More et al. [5] state that the hearing people never try to learn the sign language. And because of that deaf people cannot interact with the normal people without a sign language interpreter and eventually the deaf people get isolated in the society. Also having a personal translator for the communication can be costly. Therefore it is important to make a system that automatically recognize sign language.

**Sign Language**

As mentioned by P. Fernando el at. [3] currently there are nearly hundred sign languages can be identified around the world. American Sign Language (ASL), British Sign Language (BSL) Mexican Sign Language (LSM), French Sign Language (LSF), Italian Sign Language (LIS) and Spanish Sign Language (LSE) are just among few of them.

According to S. P. More et al. [5], Sign language is a method which uses manual communication and body language instead of acoustically conveyed sound patterns, to convey meaning.

According to V. Padmanabhan et al. [2], the dumb have their own manual-visual language referred to as sign language. Sign Language is also a non-verbal form of intercourse that's found among deaf communities

According to (11), one important means of communication method for the hearing impaired community is the use of sign language

**Sinhala Sign Language**

According to P. Fernando et al. [6] Sri Lankan Sign Language was fully built on the foundation of British Sign Language (BSL). However Sri Lankan Sign Language made lots of variations to the British Sign Language and currently consists of more than 2000 sign based words [8].

**Alphabet based sign language**

The systems that are translating Sinhala sign languages translate words and phrases into text only. There are no systems that are translating Sinhala sign language for Sinhala letters into Sinhala letter texts. Current systems are only capable of identifying few number of words and sentences among 2000+ Sinhala sign language. Alphabet of Sinhala language only consist of 60 letters. Since letters are the building blocks of any language, converting signs based on letters to text can increase the horizons to identify and translate almost any word.

M. R. Jagadish et al [6] stated that the system of communication called "finger spelling" involves spelling out words in an alphabetical language by using the letters of the manual alphabet -with hand shapes and positions corresponding to each letter of the written alphabet Conversations can be entirely finger spelled. But among deaf individuals, finger spelling is more often used in conjunction with sign language for proper names and terms for which there are no signs.

**Real time translators of the gesture recognition system**

Every researches are mainly focusing on building a real time translation of Sinhala Sign Language.

N. Kulaveerasingam et al. [7] mentioned that their research is to create a real time translator for SSL and the real time language translator leads to build effective communication between deaf people and general people.

According to P. Fernando et al. [3] their research focuses on an approach for a real-time translation from Sri Lankan sign language to Sinhalese language which will bridge the communication gap between deaf and ordinary communities.

**Categories of Gesture recognition system**

(9) say that there are basically two types of approaches for hand gesture recognition, vision-based approaches and data glove approaches. Also According to S. P. More el at. [5] the gesture recognition is mainly categorized into vision-based approach and Haptic-based approach. The vision based approach captures the movement of the signer’s hand using cameras. The haptic-based approach deals with instrumented gloves affixed with measurement devices which track hand movements. Furthermore they said that in vision-based approach, the cameras record the ever-changing image and position of the hand because the user signs and also the pictures are then processed to retrieve the hand form, position and orientation. Also as mentioned by them, the haptic-based approach methods require more power supply and it is very costly.

According to M. R. Jagadish et al [6], they specially deals with the haptic-based approach because owing to large data and complex computation involved in vision-based approach.

As N. Kulaveerasingam et al. [7] states, the existing systems are not affordable to use on behalf of impaired people. The economy is not flexible to implement such a system using the new technologies, especially in an Asian country like Sri Lanka. So their research is done based on vision-based approach.

**Glove methods (device based/ Haptic-based ) approach to convert hand gestures into required output**

The application model of M. R. Jagadish et al. [6] is a device that translates sign language of deaf-mute person to synthesized text and voice for communication. Their methodology provides a map for developing a digital wireless data glove which is fitted with Flex sensors. Flex sensors are analog resistors that function as analog voltage dividers which are sense the gestures of a person in the form of bend of fingers and tilt of the hand fist. They used flex because it is ideal for any application or device that requires the measurement of a repetitive bending, striking deflection, acceleration or range of motion. Additionally, Flex is proven for high speed impact measurements. In the proposed system they are implementing the FLEX sensor. The input data glove detects the hand gesture done by the deaf-dumb person wearing it. Next provides the analog input to the microcontroller for further interpretation according to the database. The final output is observed on the LCD display and the speaker.

V. Padmanabhan et al. [2] proposed a new technique called artificial speaking mouth for dumb people. In order to overcome the complexity of understanding the way of conveying the messages, the artificial mouth is introduced for the dumb peoples. According to dumb people, for every motion they have a meaning. That message is kept in a database. Likewise all templates are kept in the database. In the real time, the template database is fed into a microcontroller. This system is based on the motion sensor. The motion sensor is fixed in their hand. For every action the motion sensors get accelerated and give the signal to the microcontroller. The microcontroller matches the motion with the database and produces the speech signal. The system also includes a text to speech conversion (TTS) block that interprets the matched gestures. The output of the system is using the speaker.

The research paper uses database to keep meaning of the motion and fed into a microcontroller. Also fixed a motion sensor in hand and for relevant action give signals to microcontroller. Microcontroller compares those with database. They use a text to speech conversion and speaker to output the action.

**Vision-based approach to convert hand gestures into required output**

Because there are drawbacks in the haptic-based approach, researches of the gesture recognition of sign language has done by using vision-based approach.

According to S. P. More el at. [5], to overcome the drawbacks of the Haptic-based methods, they use the static hand gesture recognition system using digital image processing which is a vision based approach

H. C. M. Herath [4] presents a low cost approach to develop an image processing based Sinhala sign language recognition application for real time applications.

As mentioned in N. Kulaveerasingam et al.[7] while most of the projects use high technical features with high cost, “Nihanda Ridma” going to be an economical and affordable project, since it uses the vision based approach and the low cost technical equipment.

(9) says that The reason for choosing a system based on vision relates to the fact that it provides a simpler and more intuitive way of communication between a human and a computer.

**Methodologies used for image preprocessing**

The model of S. P. More et al. [5] will be used to recognized hand gesture captured from webcam. Before building the model, images will be captured for each hand gesture, which are the fist, index, palm, and little fingers, for different people, scales, and rotations and under different illuminations. The background has no texture or objects (white wall). Then it guarantee that all the key points extracted from training images using the SIFT algorithm will represent the hand gesture only.

In the prototype developed for the project H. C. M. Herath [4], a green background is used to capture the image for the simplicity of the implementation. And also used Matlab simulation package and a portable camera (Intex Model No IT-309WC, 16MP). First, the RGB image captured from the web camera is separated into the three matrices, red (R), green (G) and blue (B). Next G matrix is subtracted from the R matrix. This is done because it was experimentally found that red is the most dominant color of the skin as shown in the figure 1 and the background used is in green color as well. Shadows are removed in this process as in figure 2. In figure 2 (b) shadow effect remove by subtracting G matrix from R matrix and it is convert to binary image as in figure 2 (c Then again the resulted image is converted to binary image by defining a threshold. This is generated to facilitate faster mapping. The resulted binary image accuracy is depended on lighting condition at which the image is captured. If the lighting intensity is sufficient to capture the image with its natural colors or closer to natural colors then the binary image is noise free as illustrate in figure 4. This is how the image preprocessing done in H. C. M. Herath [4].



Figure 1:R, G, B value of a point in the hand



Figure 2:Shadow effect removal

**Methodologies used in image processing**

S. P. More et al [5] proposed that the feature extraction was done efficiently using SIFT computer vision algorithm. They design the SIFT algorithm for hand gesture feature vector. According to them SIFT features are distinctive and invariant features extracted from images that allow for efficient matching with various other viewpoints of the extracted features that may exist in the same or different gestures. The features are invariant to image translation, scaling, rotation and partially invariant to illumination changes. The SIFT features described in their implementation have been computed at the edges which are invariant to scaling, rotation, addition of noise. The computation of SIFT image features is in four basic steps. They are scale-space local extreme detection, key point localization, orientation assignment, key point descriptor. The advantage of using the algorithm is high processing speed which can produce results in real time.

After preprocessing the images H. C. M. Herath [4] proposed following method to process the images. After preprocessing images, the boundaries of the hand are identified by drawing smallest possible rectangle around the hand and the image is cropped to extract the region to interest. Then the cropped image is equally divided in to four parts as shown in Figure 5(b). Next centroid of each segment is calculated as shown in Figure 5 (c). The (Height/y) and (Width/x) ratios are calculated for each segment and then they are compared against precalculated values that are in the database. Errors of the ratios are also calculated by subtracting the ratio calculated for the real time image from the ratio calculated for the real time image. Finally, the image with minimum error is selected as the matched image.



Figure 3:Segmented hand and ratio calculation

As A.-A. Bhuiyan [1] proposed the ASL The ASL recognition system comprises with two segments. The feature extraction and the identification of signs. The feature extraction process is initiated with an image processing procedure, which involves an algorithm to detect and segment various desired segments of the sign.

First each color image is resized. Images of signs were resized to 30×24, by default uses nearest neighbor interpolation to determine the values of pixels in the output image. This research employs a lowpass filter before interpolation to reduce aliasing. Next step is that the hand images may be of poor contrast because of the limitations of the illumination conditions. the proposed fuzzy histogram equalization algorithm is used to reimburse for the illumination conditions and recover the contrast of the image, as shown in Figure 4. The ASL images are sometimes corrupted by numerous sources of noise. Therefore, Prewitt filtering is used to suppress the noise in the next step.



Figure 4:Fuzzy Histogram equalization

In the next step skin color segmentation is organized with visual information of the hand skin colors extracted from different images. This research uses HSV color space for skin color segmentation. In the HSV color space a color is described by three attributes: hue - the visual attribute of color sensation linked with the dominant colors, saturation -implies the relative purity of the color content and value- measures the brightness of a color. Since the human skin colors are clustered in color space and differ from person to person and of races, so in order to detect the hand parts in an image, the skin pixels are thresholded empirically. In this research, the hue values are chosen as h= [0, 40]. The detection of hand region boundaries by such an HSV segmentation process is illustrated in Figure 5. The exact location of the hand is then determined from the image with largest connected region of skin-colored pixels.



1. Original Color Image
2. Color Segmentation
3. Connected Component Analysis

Figure 5:Skin color segmentation

Finally the classification phase involves neural network training for the recognition of binary image patterns of the hand. In the classification stage, an Adaptive Resonance Theory (ART) neural network is employed. The ART contains 30×24 neurons in the input layer, 604 (70% of input) neurons in the hidden layer, and 26 neurons in the output layer.

**Methodologies used when using a Kinect camera to capture the gesture**

System of N. Kulaveerasingam et al. [7] had used Microsoft Kinect with AForge.NET Framework and the System has a multimedia database to store data using Microsoft SQL Server 2008. SQL server has the ability to store images, diagrams, graphical animations, sound and moving pictures as Binary Large Objects (BLOBs). Multimedia data have sorted in a suitable DBMS in a standardized and integrated manner.

The system of N. Kulaveerasingam et al. [7] is capable of capturing gestures one by one. It does not able to capture continuous gestures. Figure 6 enable user to record save gestures of deaf person one by one. User can compare the input gesture with default black screen by clicking the Compare button. Compare button navigates to Figure 6. Figure. 7 Show the comparison interface which compare hand gesture of deaf person with default black screen. Comparison done on the pixels of both images. The relevant word is displays if the comparison returns true. Error message is displays when the pixels are mismatched. Figure 8 Show the comparison interface which compare hand gesture of deaf person with default black screen. Comparison done by comparing pixels of both images. The relevant word is displays if the comparison returns true.



Figure 6:Gesture recognition Interface



Figure 8:Compare images



Figure 8:Comparison interface

Main objective of the proposed research of P. Fernando et al [3] is to develop software-based prototype, which can translate Sri Lankan Sign language into Sinhalese language. Microsoft Kinect SDK version 1.8, Microsoft Visual Studio 2010 IDE used to build the software using visual C# programming language. Furthermore, Windows Presentation Foundation (WPF) used since it was highly supported for representing visual information.

Each frame of the video that have when capturing the gesture will be retrieved by using kinect XBOX 360 device in order to track hand movements. As the initial point of implementation skeleton data stream and color data stream of Kinect sensor will be used. Once isolated frames are retrieved and analysed, information will be redirected to the gesture preprocessing module.

Main functionality of this module is to extract the feature points, which needs to be used for gesture detection. In gesture preprocessing by removing unnecessary details such as background information and unwanted skeleton points and using the useful skeleton points generate the feature frame. System implements a two-step normalization process done for extracted skeleton joint coordinates, before further processing. As the first step of normalization, a new center point will be calculated based on the shoulder coordinates of each user. Once the center point is repositioned, all observed coordinates will be aligned according to the new center point to make sure that the final output does not depend on the position of user while performing the gesture. As the second step of process, normalization for each coordinates will be performed based on length of two shoulder points, under the assumption of skeleton points which are symmetrical to ensure that the produced output does not depend on the physical size of the user. After the normalization process, candidate gesture identifier frame is generated. This process will be repeated until system receives minimum number of data frames which needs to detect the gesture.

They use the Training mode and store sufficient number of gesture identifier details within the gesture dictionary database. translation mode can be used and gesture identifier data will be sent to the dictionary database in order to perform a comparison and recognize the gesture. Each gesture will be performed under a window of 32 frames. From each frame, 3D coordinates of each skeleton point will be extracted and normalized. Each normalized coordinate data will be stored in the dictionary file, grouped according to sign word. Individual data files are generated for each gesture and combination of all data files form a single gesture dictionary. During the initial stage of development, five training samples have been used per gesture and dictionary of fifteen sign words has been used in the system. Therefore, final gesture dictionary consists of seventy five trained samples altogether. Once the dictionary is generated, system uses the classification module for comparing user performed signs and the training sample. Data which has the highest matching sequence will be selected as the gesture.

proposed research designed and implemented two step gesture Identification algorithm, where step 01 is based on Dynamic Time Warping algorithm and step 02 is based on Nearest Neighbor classification. Since gesture matching process follows the comparison between two sequences (Real time coordinate data and pre-trained sample data), Dynamic Time Warping algorithm (DTW) will be used with enhancements. In addition, nearest neighbor classifier will be used to choose the best matching gesture name based on the DTW classification results.

**References**

[1] A.-A. Bhuiyan, “Recognition of ASL for Human-robot Interaction,” p. 6, 2017.

[2] V. Padmanabhan and M. Sornalatha, “Hand gesture recognition and voice conversion system for dumb people,” vol. 5, no. 5, p. 5, 2014.

[3] P. Fernando and P. Wimalaratne, “Sign Language Translation Approach to Sinhalese Language,” *GSTF Journal on Computing (JoC)*, vol. 5, no. 1, 2016.

[4] H. C. M. Herath, “IMAGE BASED SIGN LANGUAGE RECOGNITION SYSTEM FOR SINHALA SIGN LANGUAGE,” p. 5, 2013.

[5] S. P. More and A. Sattar, “HAND GESTURE RECOGNITION SYSTEM FOR DUMB PEOPLE,” *International Journal Of Engineering*, vol. 3, no. 2, p. 4.

[6] M. R. Jagadish, R. Gayathri, R. Mohanapriya, R. Kalaivani, and S. Keerthana, “Hand Gesture Recognition System for Deaf and Dumb Persons,” vol. 2, no. 1, p. 8, 2018.

[7] N. Kulaveerasingam, S. Wellage, H. M. P. Samarawickrama, W. M. C. Perera, and J. Yasas, “‘The Rhythm of Silence’ - Gesture Based Intercommunication Platform for Hearing-impaired People (Nihanda Ridma),” Dec. 2014.

**(8) Sign Language Recognition using Convolutional Neural Networks**

**(9)** American Sign Language Character Recognition Using Convolution Neural Network

(10) **Real-time American Sign Language Recognition with Convolutional Neural Networks**

**(11) Real-time sign language recognition based on neural network architecture**

**Use CNN for image classification**

The system that (8) has implemented use the Microsoft Kinect, convolutional neural networks (CNNs) and GPU acceleration. The system of them able to recognize 20 Italian gestures cross-validation accuracy of 91.7%.

According to (9), their main focus of the work is to create a vision based system to identify Finger spelled letters of American Sign Language. They use CNN to implement their system.

(10) produced a robust model that consistently classifies letters a-e correctly with first-time users and another that correctly classifies letters a-k in a majority of cases. They use Convolutional Neural Networks in real time to translate a video of a user’s ASL signs into text.

(11) Their architecture is being proposed using the neural networks identification and tracking to translate the sign language to a voice/text format. They use CNN model to recognize letters in American Sign Language. First they will get a video sequence of the signer as the input from camera.

**Dataset**

(8) used the data set from the ChaLearn Looking at People 2014 [5] (CLAP14) challenge. They used 6600 gestures in the development set of CLAP14 for their experiments, 4600 for the training set and 2000 for the validation set. The test set of CLAP14 is also considered as the test set for this work and consists of 3543 samples.

(9) used an image dataset consists of ASL gestures from [1]. The dataset consists of 2524 depth images with 70 images per category. Each category represented a different character of ASL. This dataset was then augmented to create a dataset of 14781 images. Out of this dataset, 75% images were used for training and remaining 25% images were used for testing.

(10) Uses color images. They are close-ups of hands that span the majority of the image surface. They utilize ILSVRC2012 dataset, the Surrey University and Massey University ASL datasets in order to apply transfer learning to their task. Since there was little to no variation between the images for the same class of each signer, they separated the datasets into training and validation by volunteer. Four of the five volunteers from each dataset were used to train, and the remaining volunteer from each was used to validate.

**Preprocessing**

According to (8) first they have crop the highest hand and the upper body using the given joint information. The preprocessing results in four video samples (hand and body with depth and gray-scale) of resolution 64x64x32 (32 frames of size 64x64). Furthermore, the noise in the depth maps is reduced with thresholding, background removal using the user index, and median filltering.

According to (9), they have read and resize each of the image to the similar size of 224x224 pixel. As for them Only when all of the images in the dataset are of the same size can the images be fed into a neural network for training. The mean value of RGB over all pixels was subtracted from each pixel value. The mean is subracted because the model involves multiplying weights and adding biases to the initial inputs to cause activations then backpropagated [10] with the gradients to train the model. It is important that each feature has a similar range, in order to prevent the gradients from getting out of control.

(10) their Both datasets contain images with unequal heights and weights. Hence, they resize them to 256x256 and take random crops of 224x224 to match the expected input of the GoogLeNet. Also they have zero-center the data by subtracting the mean image from ILSRVC 2012. Since the possible values in the image tensors only span 0-255.

(11) states that, in order to satisfy the memory requirements and the environmental scene conditions, preprocessing of the raw video content is highly important [14]. They use a moving average or median filter as filteration preprocess. Next they did the Background subtraction. Because this is Neural network. They did feature extraction. Introduction of Point of Interest (POI) where The state of the hand gestures are given by the attributes called Point of Interest (POI) of the hands. The feature vector consists of 55 features.

**Augmentation**

According to (9) they augmented to produce several images from each image, thus increasing the size of the dataset and also tackling the problem of overfitting. Maximum ranges or degrees for shear, zoom, horizontal and vertical shifting were specified in the model

(10) They make horizontal flips of the images since signs can be performed with either the left of the right hand, and the datasets have examples of both cases. By this method they have augment their dataset

**Proposed architecture**

According to (8), For the pooling method, they use max-pooling in their ayatem. In max pooling only the maximum value in a local neighborhood of the feature map remains. They Use 2D convolutions, because it resulted in a better validation accuracy than 3D convolutions. The architecture of their model consists of two CNNs with Each CNN is three layers deep. Finally it have a classical ANN with one hidden layer. Also, local contrast normalization (LCN) as in [10] is applied in the first two layers and all artificial neurons are rectified linear units (ReLUs [14], [6])

According to (9), their Model that Used for training the dataset was inspired from VGG16 model. So VGG 16 model which is a deep convolutional neural network model used as the architecture of their CNN model. The max pooling layer used as the pooling layer. ReLU used as the activation function.

(10) They utilize a pre-trained GoogLeNet architecture trained on the ILSVRC2012 dataset, as well as the Surrey University and Massey University ASL datasets

(11) They used a new scheme called combinational neural networks (CNN). A three layer network called back propagation is used to build the CNN.

**Generalization and Training**

According to (8) dropout and data augmentation are used as main approaches to reduce overfitting during the training of the application.

**Result**

The system of (8) observed a validation accuracy of 91.70%. The accuracy on the test set is 95.68% and they observe a 4.13% false positive rate, caused by the noise movements.

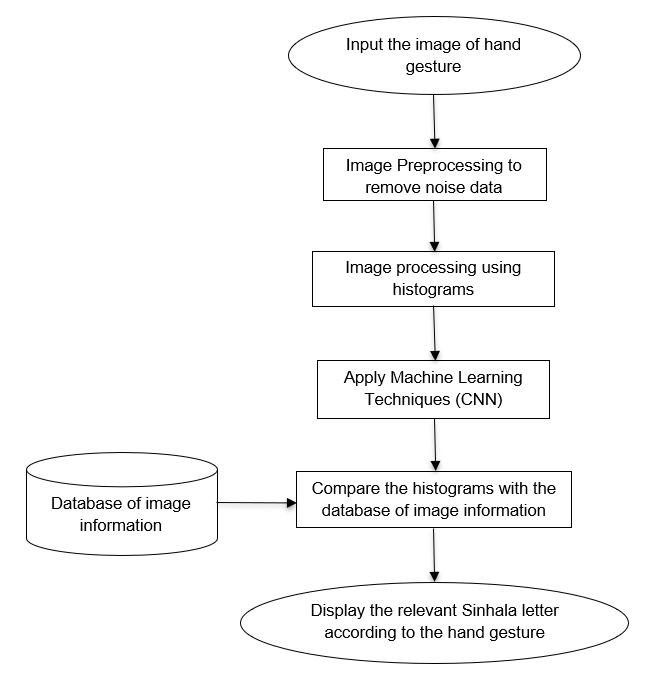
The accuracy of the model of the system of (9) obtained using Convolution Neural Network was 96%.

(10) They evaluate two metrics in order to compare their results with those of other papers. The most popular criterion in the literature is accuracy in the validation set which is the percentage of correctly classified examples. One other popular metric is top-5 accuracy, which is the percentage of classifications where the correct label appears in the 5 classes with the highest scores. Additionally, we use a confusion matrix, which is a specific table layout that allows visualization of the performance of the classification model by class.

(11) This sign language recognition approach requires a computer with at least 1GHz processor and at least 256 MB of free RAM. The training set consists of all alphabets A to Z (26 patterns).

**1.0 Methodology**

In this chapter, a cnn model to detect hand gestures for a Sinhala letter is presented. This procedure is initiated using a preprocess method for hand gesture images. Then the preprocessed images are used in the feature extraction process in CNN model. To do that the preprocessed image is going through some processing steps inside the CNN model. Finally, the model will predict the sinhala letter for corresponding hand gesture by using a desktop application which is connected to the server



**Image Processing (What is image processing)**

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies.

it is a use of computer algorithms, in order to get enhanced image either to extract some useful information.

There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers.

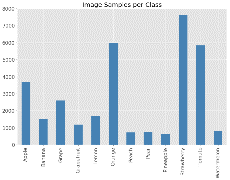
Some of the image processing techniques are pre-processing, enhancement, morpho

The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

Here we are using digital image processing in this research

**Dataset**

**(image dataset per class kiyala graph 1k add karanna puluwan)**



First, we need to collect our data and put it in a form the network can train on. This involves collecting images and correctly labeling them. This dataset is originally obtained and produced for purpose of this research. In this research as a limit we only detecting static hand gestures by using our model.

In Sinhala alphabet, there are 60 letters. Among those, 26 letters have static hand gestures while other letters have dynamic hand gestures. So we are only consider those 26 sinhala letters which are having static hand gestures. So there are 26 categories of letters we have to recognize by the CNN model. Also we are only taking the right hand of the people for take images. Because by the rotation function which we have in the model we are implementing this application, the image will give the mirror images. Then it will become the images of the left hand.

**We use color images for the dataset in this research. They are close-ups of hands that span the majority of the image surface.** (meka research paper 1ka wakyak). We used a green color for the background of the image which the hand held against a green color background. Because the furthest color from the skin color is green. Some of the images have a considerable difference with other images while some hand gestures of images have slightly difference with other hand gesture images

How dataset separate?

The dataset is divided into two categories. One is training set which are we going to use to train our model. The other category is test set which we are going to test our model and take decisions (e results matha). As further describe of the training set and test set of our dataset, there are 10 images have in one category. So there are total number of 260 images in the dataset. And there are 3 images in one category in our dataset. So there are total number of 78 images in our test set. As a result of that, the whole dataset contains 338 images. If this describe with persentages, there are 75% of training images have in our dataset. And there are 25% of test set images have from the whole dataset. This partitioning ratio is an important aspect. if number of points in the whole data set is large then any division may work fine but this data set is limited, division ratio may play a crucial role. So in this research we use the same split ratio used as in the sklearn package

**Preprocessing**

Preprocessing is used to prepare our dataset before fed it into the CNN model to train. **Preprocessing** refers to all the transformations on the raw data before it is fed to the machine learning or deep learning algorithm. For instance, training a convolutional neural network on raw images will probably lead to bad classification performances ([Pal & Sudeep, 2016](https://ieeexplore.ieee.org/document/7808140/)). The preprocessing is also important to speed up training (for instance, centering and scaling techniques, see [Lecun et al., 2012; see 4.3](http://yann.lecun.com/exdb/publis/pdf/lecun-98b.pdf)). - <https://www.freecodecamp.org/news/https-medium-com-hadrienj-preprocessing-for-deep-learning-9e2b9c75165c/>

In this research we are using the convolutional neural network to categorize the images into 26 hand gesture categories. For a CNN model, it needs a very little preprocessing. Also data do not get truly *pre*-processed in Keras, Instead the image preprocessing is doing by the model itself when the images are training. So we don’t need to worry about to apply image preprocessing techniques like *(****mean normalization****,****standardization****, and****whitening)***------------------- for the image dataset.

We wouldn’t be able to say what type of preprocessing of image is needed before feeding it into the Deep Network for feature learning. the suitable preprocessing depends on our problem domain.

So if we consider our research problem domain, all the images are taken under same parameters like background color, same side of the hand in this research. So we don’t need to apply lots of images preprocessing techniques. Therefore we have applied only two suitable image preprocessing techniques.

We have rescale our images. All images are same rectangular size shaped. it’s time to scale each image appropriately. We’ve decided to have images with width and height of 255 pixels. We’ll need to scale the width and height of each image by a factor of 1./255. There are a wide variety of up-scaling and down-scaling techniques and we usually use a library function to do this for us.

Also we did resizing. The original image is 2448 \* 3264. We have resized it into 128 \* 128. **we do resize our image during the pre-processing phase because** some images captured by a camera and fed to our CNN model vary in size, therefore, we should establish a base size for all images fed into our CNN model

**Proposed Architecture**

Uses Convolutional Neural Networks (CNN) in real time to translate a video of a user’s SSL signs into the letter. This is done by 3 steps

1. Obtaining video of the user signing (input)

2. Classifying each frame in the video to a letter

3. Reconstructing and displaying the most likely letter from classification scores (output)

CNN

Image processing can be done

With the dawn of a new era of A.I., machine learning, and robotics, its time for the machines to perform tasks characteristic of human intelligence. Machines use their own senses to do things like planning, pattern recognizing, understanding natural language, learning and solving problems. And Image Recognition is one of its senses!!!

From Automated self-driven cars to Boosting augmented reality applications and gaming, from Image and Face Recognition on Social Networks to Its application in various Medical fields, Image Recognition has emerged as a powerful tool and has become a vital for many upcoming inventions.

Nowadays we can create our own Image Recognition Classifier with a few lines of code, thanks to the modern day machine learning libraries.

Used for image classification and recognition

Network learns to extract features automatically while training

Advantages

Easy to train

Have fewer parameters than fully connected networks with the same number of hidden neurons

The Keras library in Python makes it pretty simple to build a CNN.

Convolutional 2D

Max Pooling

Activation layer

Fully Connected Layer

**Data Augmentation**

* **Increase the amount of training data using our original training data**
* **Prevent the classifier from overfitting[1][3]**
* **Augmentation methods**
  + **Shear shifting**
  + **Zoom shifting**

**Evaluation Metric**

**Implementation**