GSTF Journal on Computing (JOC)
DOI: 10.5176/2251-3043_4.4.353
ISSN:2251 - 3043 ; Volume 5, Issue 1; 2016 pp. 52-59

Sign Language Translation Approach to Sinhalese Language

Pumudu Fernando,
Informatics Institute of Technology, Sri Lanka

Dr. Prasad Wimalaratne
University of Colombo, Sri Lanka

Abstract - **Sign language is used for communication between deaf persons while Sinhalese language is used by normal hearing persons whose first language is Sinhalese in Sri Lanka. This 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. This study further focuses on a novel methodology of enabling distance communication between deaf and ordinary persons. Once the sign based gestures captured by depth sensing camera, series of feature extraction techniques will be used to identify essential attributes in gesture frame. Identified feature frame will be compared with pre-trained gesture dictionary based on classification techniques, in order to identify gesture based word. Detected word will be displayed for ordinary user or could be used for communication between two individuals in two different geographic locations. Proposed prototype has provided an overall recognition rate of 94.2% for a dictionary of fifteen signs in Sri Lankan sign language.**

***Keywords*—gesture recognition, Sign Language, Sinhalese**

I. INTRODUCTION

Spoken language has become the most popular and useful communication media among humans regardless of the country or the society. Unfortunately, not every human is capable of using spoken languages due to hearing impairment or inability to speak. There are nearly 360 million people among the world population who are reported as hearing impaired, and this is 5% of the total population [1]. Such persons use Sign language as their main communication media.

Focusing on Sri Lankan context, Sinhalese is the native language used by 74.9% of the total population [2] where as Sri Lankan sign language is used by deaf and mute community. Since most of the normal hearing (NH) people do not understand sign language, deaf persons tend to find it difficult in getting their day today work done especially related to public services such as banks, hospitals and police stations, etc. Therefore, communication becomes a huge barrier between deaf and ordinary persons, unless there is an interpreter available in these places. Due to this problem, persons who use sign language, become isolated within the society and may feel overwhelmed about their lives.

Another main problem deaf persons face is inability to communicate with others who are at a geographically different location. However, normal hearing people of this position have different ways of interacting with others using Social Networking and variety of chat applications [3] even though deaf people are deprived of these options. Tools like Skype, allow communication between deaf or mute persons who uses sign language since it contains visual communication channel. But such tools could not be used for communication between normal hearing and deaf persons since no translation facilities are provided.

Main objective of this research is to develop an approach to address above problems, and validate the approach by designing and implementing a real time prototype system which is capable of translating selected set of signs from Sri Lankan Sign language into corresponding Sinhalese words. In addition to that, proposed research suggests, implementing a hybrid method in distance communication similar to ordinary chat applications which would enable deaf people to communicate easily. Proposed method allows sending messages through sign language, Sinhalese language or both.

II. BACKGROUND

A. *Sign Languages*

Currently there are nearly hundred sign languages can be identified around the world[6]. 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. [6] Almost every spoken language has a respective sign language, developed based on above mentioned sign languages.

B. *Sri Lankan Sign Language*

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 [4]. There are nearly 25 schools built in different parts of Sri Lanka in order to provide standard education for deaf community [5]. As per the latest census carried out, there are nearly 70,000 persons who use Sri Lankan Sign language [4]

Hand Gestures, Lip Movements and Facial Expressions are used in Sri Lankan Sign language in order to convey a message. [4] [5].Combinations of above mentioned techniques are used, for certain signs, where some of the signs use one of the above techniques.

C. *Related Work*

Quan introduced [7] a Basic Sign Language Recognition system which was able to translate sequence of signs into commonly used speech language and vice versa. This sign language bidirectional translation (from signs to speech and from speech to signs) focused on the Chinese Manual Alphabet where every single sign belongs to a single letter from the alphabet [7]. The system is composed of a camera, video display terminal (VDT), speaker, microphone, and a keyboard. By using a multi-futures Support Vector Machine (SVM) classifier trained with a linear kernel in which 30 letters from Chinese manual alphabet recognized with an average accuracy of 95.55%.

Martinez Capilla [6] introduced a sign language translation system using Microsoft Kinect XBOX device as the gesture capturing device. Open NI/NITE middleware used for detection of joints of the captured frame through kinect camera. System was capable of understanding the trained gestures with 95.2% accuracy. One of the issues with this research was that there is no testing evidence of signs in a standard sign language, as the researcher has introduced his own set of signs for translation purpose.

Starner and Petland [8] invented a Visual recognition system for American Sign Language translation. They were using set of vocabulary of 40 words which consists of pronouns, verbs, nouns and adjectives. This system introduced sentence level translation of ASL and tracking camera has been used to identify users' signs. Analysing the user inputs and Sign language translation has been done using a Hidden Markov model (HMM) based algorithm. The reported accuracy rate of the system was 99.2%.

Frank Huang and Sandy Huang discovered a system to interpret American Sign Language using Microsoft kinect sensor in particular its skeletal tracker [9]. System was trained to recognize ten different signs of American Sign Language with Kinect sensor, using Microsoft skeletal tracking prototype. This research used pre trained LIBSVM as the classifier where individual gesture data has been used for training purpose.

Group of Chinese scientists introduced a new method of analysing the video frame captured through Kinect camera [10].This system introduced a new algorithm which can generate and match 3D trajectory. First, the 3D trajectory description corresponding to the input Sign language word generated by hand tracking technology provided by Kinect Windows SDK. This system initially developed to identify 239 Chinese SL words and the final implementation had two main modes named translation mode and communication mode. Under translation mode, introduced 3D trajectory matching algorithm used to identify each SL based word individually and transfer identified words for sentence recognition. Communication method used for exchanging information using SL and 3D avatar integrated to the system for simulating the sign based word. This system can be considered as a complete system created for SL translation but more testing required to explore the possibility of applying the system for other standard sign languages as well.

In Sri Lanka, there were different projects carried out based on Sri Lankan Sign Language. One such project was to implement a web-based application to provide knowledge about Sri Lankan Sign language to the public [4]. Though this system could be used as a dictionary, no interpretation services are provided.

Under the context of Sri Lankan Sign language translation, a research carried out to recognize the meaning of relevant signs using images [11]. However, there was no real time detection of gestures since this research used set of image processing techniques to detect the gestures provided as images. Therefore, hardly any research can be found which carried out a study to develop a real-time translation system for Sri Lankan sign language.

III. Design

Main objective of the proposed research is to develop a software-based prototype, which can translate Sri Lankan Sign language into Sinhalese language. Basic functionality of the system is to identify the gestures of different sign based words through a camera and process them in order to identify the respective meaning of the sign. Finally system will display the identified word in Sinhalese letters within an embedded display. This will enable normal hearing people to understand the meaning of the related sign and react accordingly.

In addition to that, proposed distance communication method allows user to send the recognized gesture through a chat application which consists of a 3d Avatar for simulating identified word using signs. Main objective of the chat application is to enable two way communication between sign language and Sinhalese language. Figure 2 provides a high level process flow of proposed prototype.

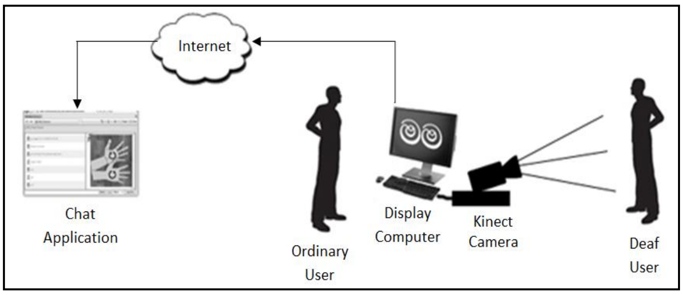


Fig. 2 High-level design of Prototype

Fifteen basic signs from Sri Lankan sign language will be used for initial identification by prototype. Mainly hand gestures will be used for representing selected signs however, the signs represented through facial expressions and finger gestures will not take into account during this research. Selected sign based words from Sri Lankan Sign Language, are provided in Table I.

TABLE I
SELECTED DICTIONARY OF SIGNS

English word	Sinhalese idea
We	අපි *(api)*
I	මම *(mama)*
Mother	අම්මා (*ammã)*
Children	ළමයි*(lamayi)*
Bus Halt	බස්නැවතුම *(bus Newathuma)*
Train	දුම්රිය *(dumriya)*
Bus	බසය *(basaya)*
Tired	මහන්සි *(mahansi)*
Come	එන්න *(enna)*
Don't	එපා *(epã)*
Go	යනවා *(yanawã)*
Where	කොහෙද *(koheda)*
What is the time	වෙලාව කීයද *(welãwa keeyada)*
Help	උදව් කරනවා *(udaw karanawã)*
Thank You	ස්තුතියි *(sthootyi)*

A. Main Process flow

As Shown in Figure 3, Deaf user starts performing the gesture in front of the camera. System will be designed with the intention of using kinect XBOX 360 device as the gesture acquisition camera. Once the camera starts recording the video, each frame of the video will be retrieved in order to track hand movements. Once isolated frames are retrieved and analysed, information will be redirected to the gesture pre-processing module.

Main functionality of this module is to extract the feature points, which needs to be used for gesture detection. Data extracted from each frame, will form a "feature frame". Prior to the translation, it is important to normalize details of the feature frame due to variations of the users who perform gestures. Different users may use the system which has different heights, varied lengths of hands, and other variations. Proper normalization must be done for the data to ensure that final output does not depend on such factors. Normalized data will be saved as gesture identifier data. This process will be repeated until system receives minimum number of data frames which needs to detect the gesture.

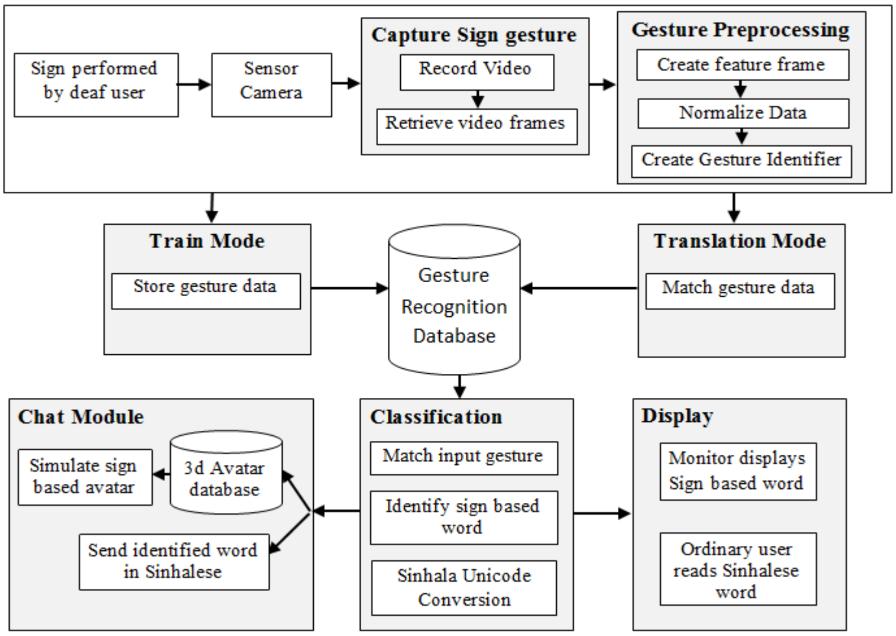


Fig. 3 System Architecture

During this stage, received visual information from the sensor will be filtered by removing unnecessary details such as background information and unwanted skeleton points. Since all points of the human skeletons will not use for performing a gesture, it is important to identify what are the joints which need to be analysed for gesture recognition. Out of the twenty joint positions, eight skeleton joints will be identified as the joints involved in performing the gesture. Therefore, joint positions shown in Table II, will be selected for generating the feature frame.

TABLE II
SELECTED SKELETONS POINTS

Left body joints	Right body joints
Left Elbow (LE)	Right Elbow (RE)
Left Hand (LH)	Right Hand (RH)
Left Wrist (LW)	Right Wrist (RW)
Left Shoulder (LS)	Right Shoulder (RS)

Generated gesture identifier data, will be used in system for two main purposes, Training and translation. System is unable to translate signs if there are no any pre-trained gesture identifier data within the system. Therefore it is a must to use the Training mode and store sufficient number of gesture identifier details within the gesture dictionary database. Once the Gesture dictionary consists of sufficient data, 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.

Classification module will be used when the system requests to perform a comparison between the received gesture identifier data and the dictionary data. This module will explore the gesture dictionary to check whether there is a suitable match in order to perform a translation. If such matching entry has found, system retrieves the related gesture name and converts the name into Sinhalese characters with the help of Unicode and displays the word in the monitor embedded to the system. If users of the system wish to send the translated information, they can use chat application integrated to the system. Chat application can be used to just send the translated word or the corresponding sign matched with the word. An integrated 3d avatar will be able to simulate the sign with the identified word. 3d models will be separated from training dictionary which performs 1:1 mapping between the identified word and the 3D simulation.

B. Specific design considerations

Storing data of the gesture identifier is one of the important aspects in system design since the gesture detection is based on data stored. System will store the data based on two aspects, Storing individual gesture data as well as all gesture data. An individual gesture data source holds the data related with the single sign (Figure 4-a). This file contains name of the sign and set of sample gesture identifier data. It is possible to have multiple samples belonging to a single gesture name. Combination of such individual gesture files will generate the final gesture dictionary as seen in Figure 4-b.

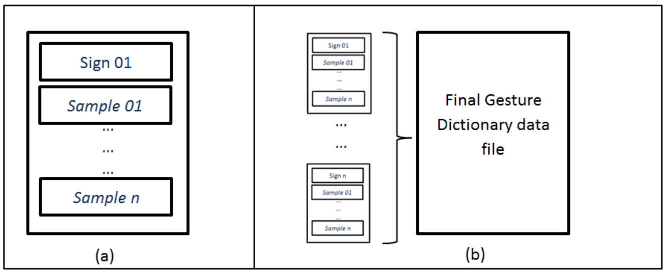


Fig.4 Data File Generation Process

Prototype architecture consists of a module for identifying the gesture. There should be a well-trained gesture dictionary as a prerequisite for the functionality of this module. Captured gesture identifier details will be used to classify the details. System will navigate throughout the gesture dictionary in order to find the most suitable matching sample for gesture data. Gesture name with highest matching probability will be selected as the gesture and the name of the gesture will be displayed in the embedded display. Figure 5 shows the flow of the design related with the classification.

Main objective of chat application module is to allow user to send the detected gesture data through network. System also allows user to access chat module without using gesture recognition module.

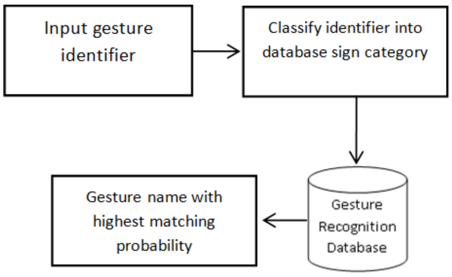

Input gesture identifier
Classify identifier into database sign category
Gesture Recognition Database
Gesture name with highest matching probability

Fig. 5 Gesture Classification Process Flow

Design of this module consists of two main sections, Character mode and Avatar mode. Character mode will be designed in a way of an ordinary chat application which can send text messages between multiple parties. In addition, Avatar mode has been designed to facilitate the communication between ordinary user and deaf user. Identified sign in the gesture identification module, will be the input for Avatar mode. System provides the facility to user for filtering recognized words they wish to send via chat application rather than sending a whole collection of words. Necessary 3D models which have the capability of simulating the gesture will be stored in a model database. Figure 6 shows the flow of chat application design.

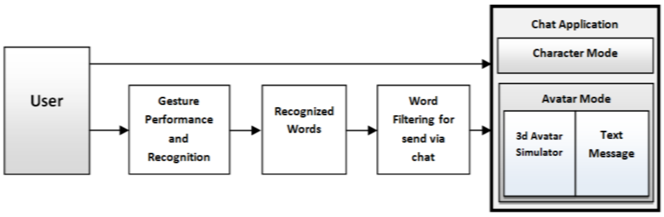


Fig. 6 Process flow of Chat Module

Prototype will display identified gesture name in Sinhalese characters. Therefore, a specific module which is capable of converting English set of characters into corresponding Sinhalese characters which has the similar pronunciation will be designed. This module has not been designed to perform any translation between English and Sinhalese words. Instead, this module takes collection of English characters as the input and converts it to Sinhalese Unicode characters [13] without changing the way of pronouncing. Basic process of this module has shown in Figure 7.

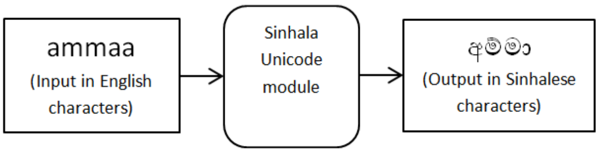

ammaa
(Input in English characters)
Sinhala Unicode module
අම්මා
(Output in Sinhalese characters)

Fig. 7 English to Sinhala Character conversion

IV. Implementation

In order to access captured data from Kinect camera, Microsoft Kinect SDK version 1.8 has been used. This SDK is supported for all the features of Kinect sensor and it provides a stable platform for developing Kinect based applications. Microsoft Visual Studio 2010 IDE will be used to build the software using visual C# programming language. Furthermore, Windows Presentation Foundation (WPF) will be used since it will be highly supported for representing visual information. In order to develop chat application, NetworkComms [12] libraries, have been used. These libraries include a dedicated set of functionalities which is capable of creating a basic chat application using C# programming language. Poser software version 6 [14] will be used to develop simulated 3d models which is supposed to integrate with chat application. Since Poser software is specifically designed for developing human character animations, it supports to create simulations of sign language based gestures using human figure.

A. Gesture Observation

As the initial point of implementation, two input data streams of kinect sensor will be used, namely skeleton data stream and color data stream. Sensor provides these data using two separate libraries, *SkeletonFrame* and *ColorFrame* in Kinect SDK[15][16]. Hence outputs provided in two different display layers will be calibrated and displayed as a single output as shown in Fig 7. Purpose of the calibration is to analyse the position of the human skeleton along with the gesture performing context.

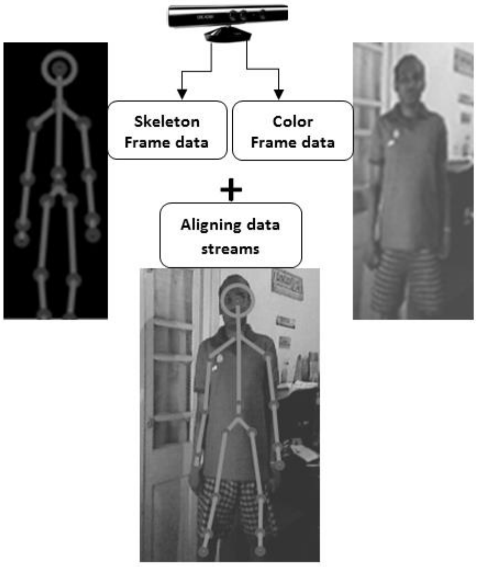


Fig. 8 Calibration process

B. Data Normalization

Proper normalization need to be done for extracted skeleton joint coordinates, before further processing. Figure 9 shows the need of normalizing the captured data.

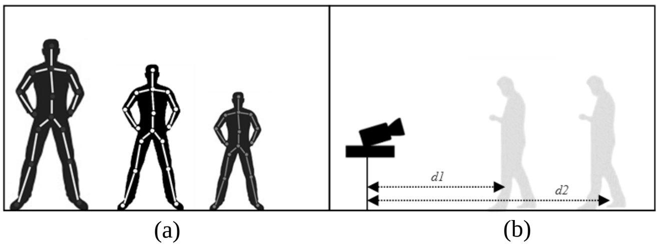


Fig. 9 normalization factors

User may be in different height (Figure 9-a) or the distance from the camera to the user may be different (Figure 9- b). In such situation, several variations of the same gesture may need to consider. This leads to redundant data storage of the system and may take long time for classification. Therefore, system implements a two-step normalization process before storing data.

As the first step of normalization, a new centre point will be calculated based on the shoulder coordinates of each user. Once the centre point is repositioned, all observed coordinates will be aligned according to the new centre point. Purpose of this step is to make sure that the final output does not depend on the position of user while performing the gesture. As shown in (1), given the set of joints,

$$J= \{LS,LH,LE,LW,RS,RH,RE,RW\} \quad (1)$$

consider a point C *x',y',z'* as the new centre point where *x', y'* and *z'* are new centre point coordinates. Assuming *x1, y1* and *z1* are the observed coordinates from Kinect sensor for Left shoulder, LS and Right shoulder RS, centre points for three coordinates are shown in (2).

$$\left.\begin{aligned} C\,x' &= (LSx1 + RSx1) / 2 \\ C\,y' &= (LSy1 + RSy1) / 2 \\ C\,z' &= (LSz1 + RSz1) / 2 \end{aligned}\right\} \quad (2)$$

Given Joint i in the joint Collection J, considering the new coordinates of J (i) are x2, y2and z2

$$\left.\begin{aligned} J(i)x2 &= J(i)x1 - C\,x' \\ J(i)y2 &= J(i)y1 - C\,y' \\ J(i)z2 &= J(i)z1 - C\,z' \end{aligned}\right\} \quad (3)$$

This alignment will be applied for all the joints in the joint collection assuming centre of the coordinate system is the middle point of users' shoulders. Figure10 shows effect of centring process. Figure 10-a represents the coordinates before centring process where figure 10-b represents the skeleton coordinate representation after centring process.

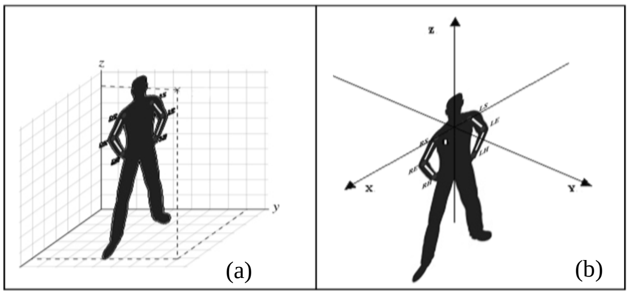


Fig. 10 Center normalization process

Once the centering process is completed, a user specific coordinate system will be created.

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. This step will ensure that the produced output does not depend on the physical size of the user. Distance between left shoulder point and right shoulder point will be calculated based on Euclidean distance of 3d coordinates. Afterward, the distance from centre point to each selected coordinate will be measured. A normalized coordinate will be calculated by dividing the joint specific distance from the distance between shoulders. This process can be illustrated through following steps.

Given two points *P(LS)*, P*(RS)* belonging to the Left shoulder and the right shoulder respectively, considering the distance,*d(s)* between *P(LS)* and P *(RS)* is shown in (4)

$$d(s) = \sqrt{(X(LS) – X(RS))^2 + (Y(LS) – Y(RS))^2 + (Z(LS) - Z(RS))^2} \quad (4)$$

Where *X(i)*, *Y(i)* and Z*(i)* are the coordinates of given point *i* Given the set of distances *D = {dLW, dRW, dLE, dRE, dLH, dRH}*, the normalized set of distances D_{norm} is obtained as shown in (5)

$$\sum_{i=1}^{n} D_{norm} = \frac{d(i)}{d(s)} \quad (5)$$

Where *n* is the number of distances from D and *d(s)* is the distance between two shoulders calculated in previous section Figure 11 shows the distances from each selected joint to the adjusted centre point.

Once the normalization process is completed, candidate gesture identifier frame is generated, which can be used for storage or comparison purpose.

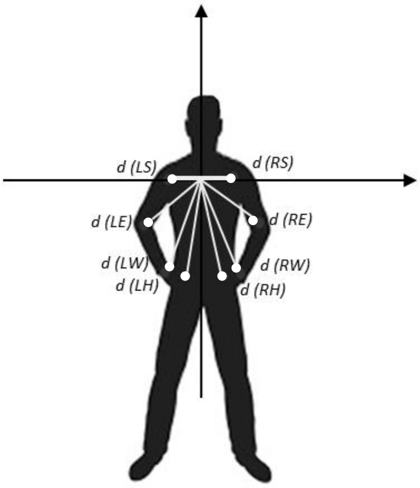


Fig. 11 Skeletal Measurements

C. Data Storage

In order to detect a particular gesture, it is important to capture desired number of frames. Therefore, each sign based gesture will be performed under a window of 32 frames. Considering each frame, 3D coordinates of each skeleton point will be extracted and normalized. Each normalized coordinate data will be stored in a dictionary file, grouped according to sign word.

Prior to translation, there should be specific amount of trained data samples, stored for comparison purpose. Hence text file based storage will be used to store the extracted coordinate information. This data consists according to the name of the gesture and the normalized set of X, Y, Z coordinates for each selected point of the skeleton. Information of thirty two such samples for a single training gesture will be available in a data file. Individual data files are generated for each gesture and combination of all data files form a single gesture dictionary.

D. Classification

Prior to classification, prototype has been trained using variety of gesture samples. 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.

For this purpose, proposed research designed and implemented two step gesture Identification algorithm (shown in Algorithm 1), where step 01 is based on Dynamic Time Warping algorithm[17] and step 02 is based on Nearest Neighbor classification[18]. 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. Figure 12 shows the expected functionality to be performed using gesture identification algorithm.

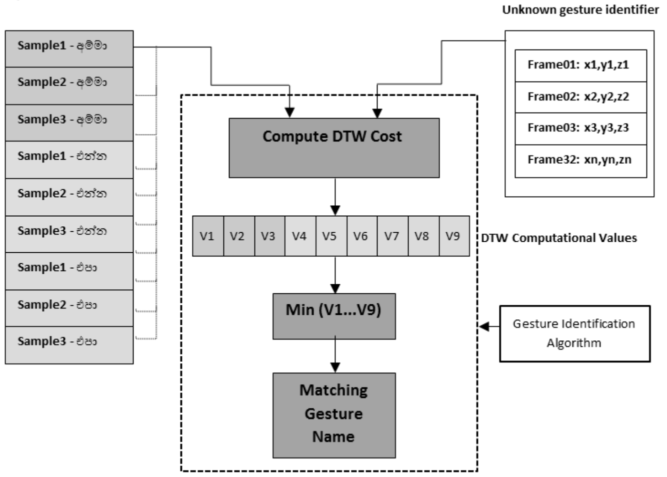


Fig.12. Flow of classification process

DTW algorithm is a time series alignment algorithm developed originally for speech recognition [17]. It aims at aligning two sequences of feature vectors by warping the time axis iteratively until an optimal match (according to a suitable metrics) between the two sequences is found [17]. It measures the similarity between two sequences based on a cost function of how much it needs to "warp" the points forward/backward in time to have them line up. This sequence alignment method is often used in time series classification.

Even though Dynamic Time Warping algorithm suits to compare similarity between two gesture trajectories, it is impossible to use the initial version of the algorithm since it has been implemented for comparing two dimensional coordinates in a given frame. This project uses 18 dimensions per frame(X, Y, Z coordinates of selected six joint positions).Therefore, Euclidian distance will be used to calculate the cost of a given two points of the inputs.

Given P and Q input arrays where each array consists of 18 dimensional data, the cost of a given frame can be calculated through following equation shown in (2).

$$\text{Cost}= \sqrt{\sum_{i=1}^{18} (P_i - Q_i)^2} \qquad (2)$$

where *i* is a coordinate point of a given joint position.

Once DTW values have been calculated input gesture need to be classified into one of the predefined gesture categories in dictionary. DTW distance based nearest neighbour classifier will be developed for this task. Calculated cost values by DTW algorithm will be sent to nearest neighbour classifier which takes generated training samples and real-time gesture samples as inputs. Once the DTW distance is calculated for each sample in the dictionary, classifier identifies the category of sign word which the real time gesture belongs, based on minimum distance between real time gesture sample and training gesture sample. Once the comparison is completed, classifier selects the gesture name which belongs to the dictionary Sample with minimum computed DTW cost.

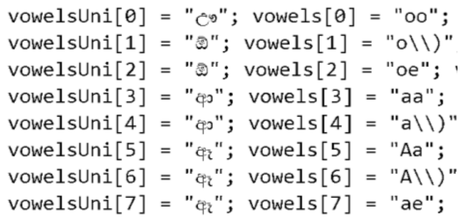
```
Algorithm 1: Gesture Identification
Step 1: Compute minimum cost
Inputs: Trained sample set, Unknown gesture sample
Output: computed cost for each trained sample
Begin
1.  Select first sample from trained sample set.
2.  Create 2D array based on sizes from trained sample and
    unknown gesture sample
3.  Select 3D coordinates from first sample of trained set
4.  Select 3D coordinates from unknown gesture sample
5.  Fill 2d array by calculating cost between each coordinate of
    trained and unknown samples based on Euclidian distance.
6.  Re-compute cost by adding current cost to Minimum cost
    from adjacent costs.
7.  Store minimum cost from 2d array related with selected
    trained sample
8.  Repeat above steps for all samples in trained set.

Step 2: Identify matching sample
Input: Generated list of minimum costs in step 1
Output: Gesture name of matching sample
1. Set first value of list as the minimum value.
2. Select next value from the list
       if (selected value < minimum value)
           Minimum =selected value
3. Repeat step two for all values in the list
4. Select gesture name with minimum value as matching
   gesture
```

E. English to Sinhalese character conversion

In order to convert English characters into Sinhalese characters, collection of regular expressions will be used. Prior to the training of the gestures, User can type the name of the gesture in English characters which will be converted into the respective Sinhalese characters which provides the similar pronunciation.

For this purpose, system uses collection of regular expressions for each Sinhalese character in the alphabet, which compares the user input and performs the conversion. Each individual character of the user input will be matched with the relevant regular expression and if a match found, particular English character will be converted into corresponding Sinhalese character compared by the regular expression. Figure 13 shows sample set of Regular expressions used in Conversion module.



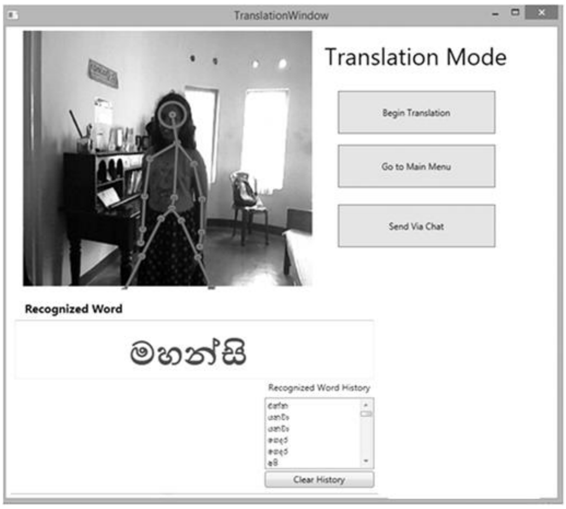
```
vowelsUni[0] = "ඌ"; vowels[0] = "oo";
vowelsUni[1] = "ඕ"; vowels[1] = "o\\)"
vowelsUni[2] = "ඕ"; vowels[2] = "oe";
vowelsUni[3] = "ආ"; vowels[3] = "aa";
vowelsUni[4] = "ආ"; vowels[4] = "a\\)"
vowelsUni[5] = "ඈ"; vowels[5] = "Aa";
vowelsUni[6] = "ඈ"; vowels[6] = "A\\)"
vowelsUni[7] = "ඈ"; vowels[7] = "ae";
```

Fig. 13 Sample regular expressions

Above mentioned modules will work together in order to translate the gesture into corresponding sign based word which displays in Sinhalese characters. Fig 14 shows the translation interface of prototype.

F. Chat module development

Main task of the chat application is to allow user to send the recognized gesture to a person at a distance. The class libraries from Networkcomms [12] are used for developing the chat application. If user wishes to send the detected gesture via chat, he/she can use the chat application module, which observes the detected gesture name/names as an input parameter. This module is capable of identifying all IP addresses connected to the network and allows user to select the desired IP address of the receiver. Each Chat Client involved with the application has a unique port number. Therefore, messages can be sent to the desired user by using the IP address along with the unique port number. All the detected gestures in gesture recognition module will be lined in a queue and sent to the chat application for retrieval and display of the necessary 3d module. Figure 16 shows a sample chat application interface.

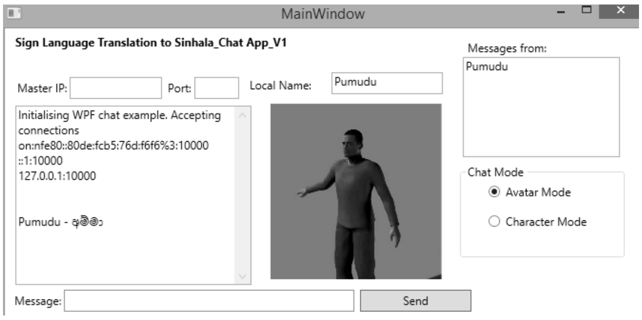


Recognized Word

මහන්සි

Fig. 14 Prototype: Translation interface

In order to develop the necessary 3D simulations of the signs to be sent, Poser software has been used. Each simulation consisted of 120 frames which displayed nearly four seconds within chat application depending on the gesture name entered by the user. Fifteen simulations has been developed for selected sign based words to be appear in chat application based on the user input. Default human model provided by the Poser software has been used for implementing sign language simulated animation. As shown in Figure 15, identified word can be sent as a series of characters or 3d simulation of the gesture.



Master IP: Port: Local Name: Pumudu

Initialising WPF chat example. Accepting connections
on:nfe80::80de:fcb5:76d:f6f6%3:10000
::1:10000
127.0.0.1:10000

Pumudu - අම්මා

Messages from:
Pumudu

Chat Mode
● Avatar Mode
○ Character Mode

Message:

Fig. 15 Prototype: Chat interface

V. Evaluation

For training purpose, fifteen selected sign based words from Sri Lankan sign language were used, where each word consists of five training samples in gesture dictionary. Hence training dictionary contains seventy five total samples taken from variety of users.

Prototype has been tested based on four main areas; accuracy of individual sign, effect of the distance from camera, effect of users height to individual gesture, and speed of performing gesture. Table III contains the number of samples used for testing each factor.

TABLE III
System Evaluation Criteria

Factor	No of Samples
Individual sign based evaluation	225
Effect of the distance from the camera	50
Users with physical height variations	50
Speed of the gesture based evaluation	50

A. Accuracy rate of individual Signs

According to the results shown in Table IV, system was able to achieve an overall accuracy rate of 92.4% out of 225 test samples. One of the common characteristic with the gestures having low accuracy rate is that those gestures consist of complex paths and both hands are used in performing the gesture. This accuracy rate was obtained by training equal number of samples for each gesture. Hence results show that some of the gestures need more training samples due to the complexity and variation of gesture pattern.

TABLE IV
Word Based Test Results

Sign	**Total samples**	**Passed**	**failed**	**Accuracy**
We (අපි)	15	12	2	80%
I (මම)	15	14	1	93.3%
Mother(අම්මා)	15	15	0	100%
Children(ළමයි)	15	15	0	100%
Bus Halt(බස් නැවතුම)	15	12	3	80%
Train(දුම්රිය)	15	13	2	86.6%
Bus(බසය)	15	12	3	80%
Tired**(මහන්සි)**	15	14	1	93.3%
Come(එන්න)	15	15	0	100%
Don’t(එපා)	15	13	2	86.6%
Go (යනවා)	15	14	1	93.3%
Where (කොහෙද)	15	14	1	93.3%
What time(වෙලාව කීයද)	15	15	0	100%
Help(උදව් කරනවා)	15	15	0	100%
Thank You(ස්තූතියි)	15	14	1	93.3%
Overall Status	**225**	**208**	**17**	**92.4%**

B. Testing based on distance from camera

Prototype was evaluated against the distance between camera and the gesture performing user. During evaluation stage five random gesture samples of the dictionary, were taken and performed from ten different ranges of the distances between camera and the user. Distance range of 0-160 (inches) was taken into account for evaluation. Figure 16 shows the distribution of the accuracy rates against the distance from the camera to the gesture performing context.

As per the results, system shows the highest accuracy rate with in the distance range between 60-120 inches with the reported highest rate of 94% (Highlighted area in Figure 16).

In between this distance range, the capturing window of the camera was able to capture the entire human skeleton. Even though the project does not take all detected skeletons into account, during the mentioned range of distance camera was more accurate in identifying and real-time drawing of skeletons. Recognition level started decreasing, once the distance range exceeded amount of 140 inches, since camera was unable to detect some of the skeleton points. However, system shown some sort of accuracy level in recognizing certain set of gestures.

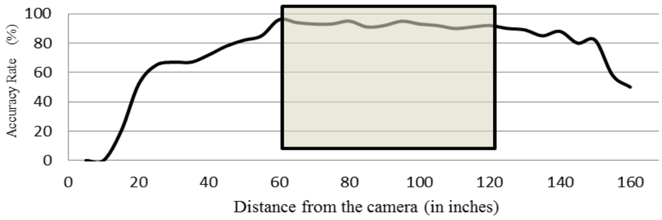


Fig. 16 Accuracy rates based on distance from the camera

C. *Users with different heights*

Five different users have been selected for this testing with different heights and fifteen random gesture samples performed by each user. One of the challenges faced during this test, was to test the system against users who were not aware of sign language gestures. Hence the success rates of the users were highly dependent on the way they perform the gesture. However, system was capable in recognizing majority of the gestures as mentioned in Table V with reported highest accuracy rate of 90%. Even though the results show that the accuracy rate is directly proportional with the height, final output was based on accuracy of individual gesture performed by the user and the amount of related samples in the training database.

TABLE V
ACCURACY BASED ON USER'S HEIGHT

Sample No	User height (in cm)	Accuracy (for 10 samples)
01	134.6	50%
02	152.4	70%
03	167.1	90%
04	172.3	80%
05	180.3	80%

D. *Accuracy based on gesture speed*

One of the main challenges of prototype implementation was inability to predict the speed of the gesture performed by the user. Therefore, the system has been evaluated using five gesture samples performed in various speeds. Gestures have been selected to cover up all complex and simple gesture paths. Since the training database contains the gestures performed up to thirty two frames, testing range of the gesture was in between 5-32 frames. Gestures were tested restricting to four main categories of frames in terms of the completion time (i.e. 5, 15, 20 and 30) Figure 18 shows the accuracy rates received for each sample against the number of frames taken to complete the gesture. According to figure, majority of gestures were not accurately identified when gesture completion frames are very low. However, sample one and two shows a high accuracy rate even with the low number of frames since the path of the gesture is not that complex and the gesture can be easily performed within a short period of time.

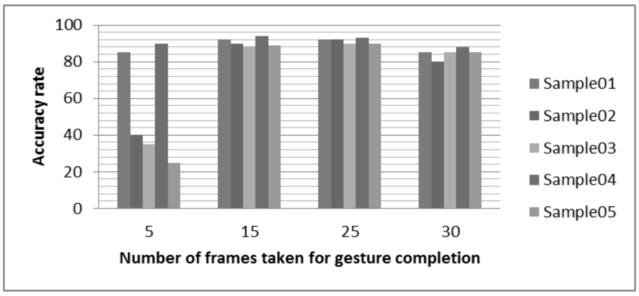


Fig. 17 Accuracy rates based on speed of the gesture

As shown in figure 17, starting from frame number 15 onwards almost all the gestures has shown high accuracy rate reflecting that provided algorithm is capable of recognizing the gesture even with varying speeds. One important observation during this evaluation phase was that recognition of some samples does not depend on recorded frames in the gesture dictionary. Therefore, it is not necessary to record the same gesture with different speeds in training phase. However, during the training stage, it is always better to train the gesture in slow frame rate because initial DTW algorithm which used for classification purpose has proved the capability of comparing the gestures with varying speeds.

E. *Evaluation Summary*

Implemented prototype has been functioning with the reported overall accuracy of 94.2% for selected individual signs under the scope of the research. Accuracy of certain gestures depends on the number of gesture samples of the training database. Gestures with complex paths needed to train with more training samples in order to achieve a relatively better accuracy rate.

Reported highest accuracy rate of 90% given for the user's height based evaluation process and the results were highly depended on the way user is performing the gesture. However results have shown that the prototype was functioning independently from the user's height.

There was an effect of the distance from the camera to the gesture performing context where the results have shown that the prototype has provided good recognition rate between the distance range of 60-120 inches from the camera to the gesture performing context. System was capable of recognizing gestures beyond the stated distance but the Kinect camera was unable to provide a clear set of skeleton information when the distance from the camera is too far.

Accuracy based on the speed of the gesture had dependability on complexity of gesture path. Gestures with simple paths were identified even with very high speed. Overall, system has taken beyond fifteen frames to provide a reasonable level of recognition.

During the evaluation stage, main focus was on testing success rate of the gesture recognition module designed, rather than testing chat module. Since chat module had a static behaviour and it had fixed set of steps to perform, chat module has shown fixed set of results in multiple iterations.

VI. CONCLUSION

Initial goal of the research was to design and implement an approach which is capable of translating Sign Language into Sinhalese language. In addition, this study has focused on the possibility of sending the translated words to a disabled or normal person at a distance with via hybrid approach. During the background study, it was clear that, there were different studies carried out to translate country specific sign languages into respective spoken languages[8][9][10]. But there were no evidence of papers regarding a study, carried out for real time translation between Sinhalese language and sign language which was the major reason for proposed research in this paper. Several phases were designed and implemented where the main focus was on accurate recognition of sign based gesture. Dynamic Time warping and nearest neighbor classification based gesture identification algorithm has been proposed. The results have shown an accuracy rate of 92.4% for detecting selected set of signs, which has proven the suitability of proposed algorithm. Additional evaluation results of the implemented prototype provided that signs performed by different users, with varying speeds can be captured and identified. Findings of this study suggested capturing window of 32 frames per gesture was reasonable for signs with simple trajectories, where the window size had to be increased for hand gestures with complex paths.

In addition, novel way of distance communication method has been proposed and implemented in this paper which has gone beyond the traditional chat application. Integrated 3d simulation with chat module allows user to send messages via words or sign based simulations.

A. Future Work

At initial stage, this research has proposed an approach to detect dictionary of fifteen sign based words in isolation which integrates only hand gestures. This can be further expanded by including more signs demonstrated through finger movements and facial expressions as well.

REFERENCES

[1] Population of sign language users. [Online].available: http://www.who.int/mediacentre/factsheets/fs300/en/

[2] Department of Census and Statistics-Sri Lanka," Census of Population and Housing - Final Report",2012,pp 129

[3] Leading social networks worldwide. [Online].available: http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[4] Sri Lankan Sign Language. Information and Communication Technology Agency Sri Lanka.[Online]. Available: http://www.lankasign.lk/index.html

[5] Adam Stone and Monika Rego, *An Introduction to Sri Lankan Sign Language.* Rohana Special School, 2007

[6] Daniel Martinez Capilla, "Sign Language Translator using Microsoft Kinect XBOX 360" , M.S. thesis, University of Tennessee at Knoxville -USA, IL, 2012.

[7] Yang Quan and Peng Jinye. "Application of improved sign language recognition and synthesis technology" in ib. In *Industrial Electronics and Applications,2008.ICIEA 2008.3rd IEEE Conference* on, pages 1629 -1634, June 2008.

[8] T. Starner and A. Pentland, "*Visual recognition of American sign language using hidden markov models,"* in: Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995.

[9] F. Huang and S. Huang, "Interpreting American Sign Language with Kinect," pp. 1–5, 2011

[10] Xiujuan Chai, Guang Li, Yushun Lin, Zhihao Xu, Yili Tang, Xilin Chen, Ming Zhou,"Sign Language Recognition and Translation with Kinect". *10th IEEE International Conference on Automatic Face and Gesture Recognition* , Apr. 22-26, 2013, Shanghai, China

[11] H. C. M. Herath, W. A. L. V. Kumari, W. A. P. B. Senevirathne, M. B. Dissanayake Image based Sign language Recognition System for Sinhala sign language. in*Proc. SAITM RSEA -2013,* April 27, 2013

[12] Creating WPF Chat application, [Online], Available : http://www.networkcomms.net/creating-a-wpf-chat-client-server-application/

[13] Samaranayake, V. K., Nandasara, S. T., Dissanayake, J. B., Weerasinghe, A.R., Wijayawardhana, H. An Introduction to UNICODE for Sinhala Characters, University Of Colombo School of Computing. 2003

[14] Poser Software for character animation, [Online], Available: http://poser.smithmicro.com/

[15] Abhijith Jana, Kinect for Windows SDK Programming Guide, PACKT publishing, 2012

[16] David Katuhe, Programming with the Kinect for Windows Software Development KIT, Microsoft Press, 2012

[17] Jo Criel and ElenaTsiporkova,"Gene Time Expression Warper: a tool for alignment, template matching and visualization of gene expression time series", Oxford Journal for Bio informatics,Vol. 22 no. 2 2006, pages 251–252

[18] Oliver Sutton, "Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction",2012

AUTHORS' PROFILE

Mr. Pumudu Fernando - graduated in 2011 in Information Technology at Faculty of Computing of Sri Lanka Institute of Information Technology, Sri Lanka. He has also completed his Masters in Computer Science from University of Colombo School of Computing. His research interests are Human gesture recognition, Image Processing, Computer Vision and Interactive Web.

Dr. Prasad Wimalaratne - PhD(Salford), BSc(Col), SMIEEE, MCS(SL) obtained Ph.D. in Virtual Environments from the University of Salford, United Kingdom in 2002. He is a senior member of IEEE. His research interests include virtual reality, computer graphics, and assistive technology. He has served as a council member of the Sri Lanka Association for the Software Industry (SLASI), vice chair of Institute of Electrical and Electronic Engineers (IEEE) Sri Lanka Section and a council member of the Computer Society of Sri Lanka (CSSL).