Introduction to Graduation Project

Ameer Alqam, Ali Nubani, Sara Yassin

Advisor: Dr. Aziz Qaroush

November 25, 2016

Abstract

This is where the abstract goes

Contents

Chapter 1

Introduction

1.1 Introduction and Motivation

Just for one man, a whole village learned the sign language and even though many of us dont deal with many hearing impaired doesnt mean Palestine doesnt have any. In fact, Palestine has % of deaf people.

Nowadays, when getting an idea, even though you think that you are probably the only one that came up with it, when you research you find out exactly how many people have had the same idea and what they did about. From our research, we found out that a good amount of people know how important it is that sign language needs to be converted to speech. We also found that many of these researchers failed to get their goal, or they simply didnt get to the needed accuracy and availability for deaf people to be able to use their product on a daily basis.

There are many experiments with the American Sign Language, German Sign Language, Indian Sign Language, and Vietnamese Sign Language, but there are almost no researches using the Arabic Sign Language.

Many designs are not as user-friendly and are not person-dependent and dont even consider the sign languages proper rules and grammar.

1.2 Problem Statement

Even though there is a percentage of deaf in our community, not many people know the sign language and so it is difficult for the deaf to communicate with anyone that doesnt know the sign language and to lead a normal life. Therefore, they need a way for everyone to understand them while still using the language they know.

1.3 Methodology

Our hope is that we solve this problem by using MEMS sensor technology. Six IMU sensors on each hand, one on each finger and one on the back of the hand attached to the microcontroller. These will hopefully be able to detect the movement of the fingers and hand and thus, capture most of the sign language gestures.

This way we are looking to capture most of the sign language gestures and will use the microcontroller alongside the multiplexer to send the data via bluetooth to a mobile phone to give the output as speech.

1.4 Report Outline

This report will firstly discuss the previous related works we studied, will talk about their technologies and disadvantages, also how our design will be different. Then, we will discuss how the system will be designed, followed by the system implementation and concluding with our results by the end of the semester and what we hope to achieve by the end of the next semester and after in the future work section.

Chapter 2

Related Works

A lot of work was made on the topic of SLR in different languages and different fields from different universities and countries. From what we studied, we classified the methodologies used for gesture recognition into two major branches based on the technology used. These branches are Video Processing and Hand Attached Sensors. Further discussion and description of those branches is in the following sections.

2.1 Video Processing

One of the major techniques used in gesture recognition (which SLR is an application of), it relies on processing a continuous stream of images (a video source). Processing is made on the video stream to determine the objects in it, the points of interest, and the shape and motion of the objects throughout the stream, from this information gestures can be distinguished from one another.

Kinect

Kinect technology is widely used in motion capture systems MOCAP. Its used in gaming, films, robotics and many other computer vision applications. The kinect is a device in a shape of a flat black box that was developed by Microsoft to be primarily used in the games that were aimed for the XBOX to recognize players movements and gestures in 3D space. It uses an RGB camera that detects objects based on their color, along with a depth sensor, which is an infrared laser projector combined with a monochrome CMOS sensor that records the video data regardless of the available colors. The recorded data is then mapped on a digital model for 3D reconstruction. The reason Kinect is adapted in a large number of startup projects is how cheap it is compared to other Rotoscoping and CGI technologies, also because of its high accuracy and speed.

https://www.microsoft.com/en-us/research/blog/kinect-sign-language-translator-part-1/ In 2012, a SLR project was started by developers at Microsoft Research Asia collaborating with the Chinese Academy of Sciences, and Beijing Union University. It uses Kinect to interpret the signs and translate them into spoken text, it can also do the reverse translation (text to signs). The project focused on the Chinese sign language since they have been already researching it for 10 years. They took data from 5 different persons for each word to establish the patterns and used machine learning and pattern recognition programming to create and recognize meanings for the signs. The project took more than a year and a half in development and it supported more than 300 of the most popular signs used in the Chinese SL. The research is still a prototype and was never available in the market.

Another good research that we found was done by the Centre of Vision Speech and Signal Processing in the University of Surrey (UK). What was interesting in this research was that they used linguistic sub-units to recognize the signs. They presented a comparison of using different sub-units, and different classification models. They made use of both the 2D and 3D data outputs from the Kinect sensors and classified them into sub-units based on the hand shape, location, motion and hand-arrangement.

When selecting their segmentation, features, and classification process for each type of processing they were very thorough with collecting the data and choosing the right features and segmentation depending on what kind of tracking they used. For example, they started based on appearance and they choose their segmentation as color segmentation and their features were selected based on location of the hands and knowledge of where the face is and based on that they choose their classification to be as follows:

Compared to other papers and researches they had data that was huge in size and that was generated properly and in the right environments. They had mainly 15 participants and the data was captured using a mobile system giving varying view points which is also a different approach than most other related works.

They achieved a rate of 99.9% on 20 signs from a multi-user data set. They also tested their work on a real environment which is challenging, and got a result of 85.1% on a user-independent 40 signs data set. In the end, the research promised future work but didnt provide any.

Leap Motion

The LeapMotion controller is a device that is used to track and recognize the hands gestures and movements using two monochromatic cameras and infrared LEDs. The software then receives the data as a stream of

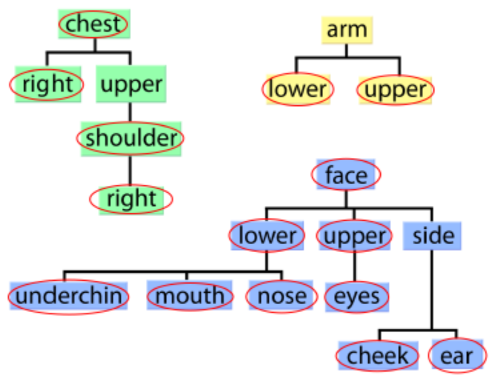

Figure 3: The three *Location* sub-unit trees used for classification. There are three separate trees, based around areas of the body which do not overlap. Areas on the leaves of the tree are sub-areas of their parent nodes. The ringed labels indicate that there are exact examples of that type in the data set.

Figure 2.1: name figure

frames from the sensor and analyzes it to construct a 3D representation by comparing the tracking data from the two cameras.

One group that used this technology was a group from Naras Womens University in Japan where they targeted the American Sign Language and used the Genetic Algorithm with the LeapMotion Controller to recognize 24 letters.

They studied previously existing methods and explained how the technology of colored gloves dyed with 6 different colors was not that wearable and a photograph picture must be taken with wearing the colored glove (not very convenient nor flexible). They also studied Kinect based recognition and showed that large space is required to obtain depth as well as the color image for the skeletal tracking, which is not easy for ordinal use. Also, they reasoned that recognizing the alphabet by finger spelling 26 letters using Leap Motion Controller would not be accurate and that the machine learning requires large computations for each new person and thus the need for efficient, flexible, accurate and simple mean for sign language recognition.

The data acquired is generated automatically by the LeapMotion API and the segmentation is achieved by finding the appropriate crossover probability and finding appropriate mutation probability. The probability calculations were repeated 10 times for each parameter. Then, the appropriate parameters obtained in are used for obtaining a sub optimal solution (a process that is performed 100 times with 1,000 individuals and 300 generations). The decision tree is automatically generated by a Genetic algorithm to obtain quasi-optimal solutions 3.4 minutes on average. The crossover probability is 0.8 and the mutation probability is 0.08.

Their results were not accurate, because they got a recognition rate of 82.71% when their data set was only 24 letters. Moreover, their system was supposed to be more user-friendly than the gloves but ended up with the same low level of efficiency since the user will have to carry a high performance computer, which is large in size, since micro-controllers cant be used for the purpose of analyzing too many frames in a second (LeapMotion generates 200 frames per second).

2.2 Hand-attached Technology

Hand-attached sensors, as we classified them, are the sensors that are directly in contact with any part of the hand whether that part was the fingers, wrist, or forearm. Many students and scholars have researched and experimented with different approaches for these sensors and in this section we examine some of the most important results and studies including AcceGlove, FlexGlove and more.

MEMS Accelerometers

A recent work that is really similar to the work we will be performing is a project of Recognizing the Vietnamese sign language using MEMS accelerometers by Duy Bui and Long Thang Nguyen in Vietnam National University, Hanoi. Basically, they used six sensors (one on each finger and one on the palm).

Their hardware included: MEMS accelerometers (six ADXL202 chips), BASIC Stamp micro-controller, and a PC. They used C++ as a language and a fuzzy rule based classification system. The sensors basically detects x-axis and y-axis points. The y-axis points toward the fingertip, providing a measure of joint flexion, while the x-axis can be used to extract information of the hands roll or yaw, or individual finger.

They experimented on 23 out of 26 letters of the Vietnamese language and got a recognition rate of 100% for 20 out of the 23 letters and 94%, 90% and 96% for the R,U and V, respectively. To get these results, they collected 200 samples and every time they receive the data from the sensing device, they first verify if the hand is at static position by comparing with previous data. They wait until the hand stops moving to start the recognition process. The pre-processed data is used to calculate the membership values the degree to which the data belongs to the fuzzy sets. They then calculate the degree to which the current data set matches each of the 22 fuzzy rules. Moreover, they enhanced the recognition process by the using the

Vietnamese spelling rules. They clarified their classification process using the following figure.

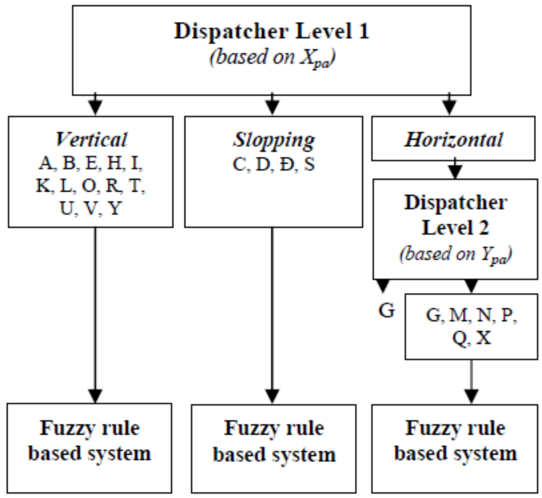

Figure 2.2: Overview of Classification Process

Their research is a pretty good research but it can only be applied to the Vietnamese sign language since it does differ from the American sign language and other sign languages. It also only focuses on the x and y axis although the sign language itself really needs the z axis in most words and therefore this is only practical for letters and numbers mostly. Another thing that could increase the accuracy of their recognition is using a gyro along with the accelerometer, and they didn't use a portable computer or a smart phone (they used a PC). Also, since they used six sensors only, they abandoned the fact that some words and expressions in the sign language need the second hand as well.

Surface Mounted EMG coupled with IMUs

Another interesting work done on SLR was a hybrid system, comprised of EMG sensors coupled with an IMU, made by Jian Wu, Lu Sun, and Roozbeh Jafari. The system used surface mounted EMG sensors on the arm to distinguish between finger positions, and an IMU on the wrist to determine the orientation and motion of the hand. The figure below shows the positions of the sensors.

The experimentation was made on 80 ASL signs, with results were 96.16% intra-subject recognition accuracy. Data was collected over bluetooth and preprocessed to reduce noise, then segmented using an average energy window while comparing it to a threshold. The threshold is weighted by a convergence parameter, and updated each time by a divergence parameter, to insure the systems handling of noise.

Features were selected from a well-established set based on previous researches, and Information Gain

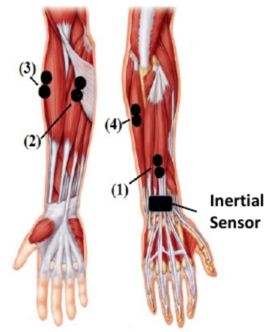

Fig. 5. Placement of sEMG electrodes.

Figure 2.3: name

was used to reduce the feature vector as well as to determine the more important features for reference. The following table shows the features selected. Then a comparison between 4 classifiers, Naive Bayes, Decision Trees, Nearest Neighbor, and SVM, all from the Weka tool was made. SVM (LibSVM with Radial Basis Function Kernel) was shown to be the best classifier with the highest accuracy. The results of the tested classifiers is plotted in the figure bellow, in reference to the number of features selected.

TABLE V. FOURTY SELECTED FEATURES

Rank #	Feature name	Rank #	Feature name	Rank #	Feature name	Rank #	Feature name
1	Mean of Acc_y	11	RMS of Gyro_x	21	RMS of sEMG1	31	Signal magnitude area of Acc_x
2	Mean of Acc_z	12	RMS of amplitude of accelerometer	22	Zero cross rate of Acc_y	32	Variance of sEMG4
3	RMS of Acc_x	13	Mean of amplitude of accelerometer	23	Variance of Gyro_z	33	Entropy of Gyro_x
4	RMS of Acc_z	14	Mean of Acc_x	24	Standard deviation Of Gyro_z	34	RMS of sEMG4
5	RMS of Acc_y	15	Signal magnitude area of Acc_x	25	Variance of Acc_y	35	Signal magnitude area of Gyro_x
6	Integration of Acc_y	16	Standard deviation of Acc_z	26	Standard deviation of Acc_y	36	Zero cross rate of Acc_z
7	Integration of Acc_x	17	Variance of Acc_z	27	Modified mean frequency of sEMG1	37	Mean absolute value of sEMG4
8	Integration of Acc_z	18	Standard deviation of Gyro_z	28	Mean absolute value of sEMG1	38	Signal magnitude area of Gyro_z
9	Entropy of Acc_x	19	Variance of Gyro_x	29	First auto-regression coefficient of Acc_x	39	RMS of sEMG2
10	RMS of Gyro_z	20	Variance of sEMG1	30	Mean absolute value of sEMG2	40	Mean of amplitude of gyroscope

Figure 2.4: name

Some important observations to be made from the experimentation, is that the features obtained from the EMG part of the system had the least Information gain among all other features gained from the IMU part. Only one of the EMG sensors gave useful information, in comparison to the IMU which had the most information gain. Another point is that the accelerometer had the highest number of features with high information gain.

The SLR models made were person-dependent, that is when the models were tested between people the

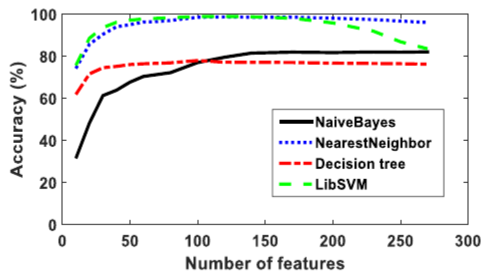

Fig. 6. Results of feature selection.

Figure 2.5: name

accuracy dramatically decreased, shown in the figure below.

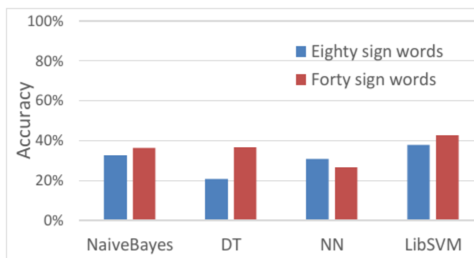

Fig. 7. Results of inter-subject testing.

Figure 2.6: name

Flex Sensor

Coming to the most recent technology in todays social media and in the technology world, the SignAloud gloves. These gloves use flex sensors that are bend measurement sensors that are used in many gloves to check for finger movement. They were developed by UW sophomores Navid Azodi and Thomas Pryor translate American Sign Language into speech and text in the University of Washington.

Each glove contains sensors that record hand position and movement and send data wirelessly via Bluetooth to a central computer. The computer looks at the gesture data through various sequential statistical regressions, similar to a neural network. If the data matches a gesture, then the associated word or phrase is spoken through a speaker.

There is a lot of work on this project that needs to be done for it to be complete and ready for commercial use. Also, the language is not proper sign language where they translated the language word for word which is incorrect. They havent published any papers so we couldnt understand their process of segmentation and classification and their actual testing results. Moreover, as mentioned by the Office of News & Information of the University of Washington While well-meaning, those students have demonstrated that they do not understand what American Sign Language is or how their act of cultural appropriation has deeply insulted the Deaf community. This project was not a collaborative project with the Deaf community, either in execution or in completion" - Lance Forshay, Senior Lecturer and Director, ASL Minor Studies Program Kristi Winter, ASL Lecturer Emily M. Bender, Professor of Linguistics.

Chapter 3

System Design

Chapter 4

System Implementation

Chapter 5

Conclusion and Future Work

Bibliography

[1] Author, "Title", Journal, Volume, Pages, Year.

This is the annotation for this bibliographic record.