American Sign Language Recognition Using an FMCW Wireless Sensor (Student Abstract)

Yuanqi Du, Nguyen Dang, Riley Wilkerson, Parth Pathak, Huzefa Rangwala, Jana Kosecka
George Mason University
Department of Computer Science
Fairfax, VA
{ydu6, ndang6, rwilker2, phpathak, rangwala, kosecka}@gmu.edu

Abstract

In today's digital world, rapid technological advancements continue to lessen the burden of tasks for individuals. Among these tasks is communication across perceived language barriers. Indeed, increased attention has been drawn to American Sign Language (ASL) recognition in recent years. Camera-based and motion detection-based methods have been researched extensively; however, there remains a divide in communication between ASL users and non-users. Therefore, this research team proposes the use of a novel wireless sensor (Frequency-Modulated Continuous-Wave Radar) to help bridge the gap in communication. In short, this device sends out signals that detect the user's body positioning in space. These signals then reflect off the body and back to the sensor, developing thousands of cloud points per second, indicating where the body is positioned in space. These cloud points can then be examined for movement over multiple consecutive time frames using a cell division algorithm, ultimately showing how the body moves through space as it completes a single gesture or sentence. At the end of the project, 95% accuracy was achieved in one-object prediction as well as 80% accuracy on cross-object prediction with 30% other objects' data introduced on 19 commonly used gestures. There are 30 samples for each gesture per person from three persons.

Introduction

ASL utilizes a large variety of hand and facial movements to communicate. It is most often used by those who are deaf or hard of hearing, though many individuals from the hearing community communicate with it as well (National Institute on Deafness and Other Communication Disorders (NIDCD) 2019). In order to improve the communication between people who are familiar with ASL and those who are unfamiliar with ASL, researchers have been working to implement a program using various types of devices to recognize the gestures. Two widely used devices are cameras (Zafrulla et al. 2011) and motion-detection sensors (Potter, Araullo, and Carter 2013). Although they have shown great success in recognizing ASL, they do have a variety of drawbacks, e.g. the camera interferes with user privacy and the IMU is invasive to the user as it requires them to attach it to their body. Instead, this research team proposes to solve this problem by using a wireless sensor (Fig. 1) which can both protect the user's privacy and eliminate the need to attach devices to the user's body. The sensor will track the movement of the user by reflected signals and develop the general shape of the user's body parts and movement as shown in Fig. 2 left. Transforming the data into a more desired representation allows the machine to learn the patterns of the gestures from the underlying data and translate ASL to English. This paper will focus on the methodology of data representation and how it could be used to perform well in this task.

Figure 1: FMCW Wireless Sensor

Related Work

Traditionally, the most common tools used in sign language recognition are camera sensors and motion-detection sensors . For cameras, individuals typically use Kinect (Zafrulla et al. 2011) allowing the device to process the images along with the depth information to let machines learn the gestures and make predictions. However, the device is very sensitive to light conditions, expensive, and presumably unaffordable for many individuals in terms of real-world application. For motion-detection sensors (Potter, Araullo, and Carter 2013), people usually embed the sensors in gloves, which are worn on the hands, or directly use IMU (Inertial Measurement Unit) attached anywhere on the body of the user which may lead to discomfort with the device usage.

Methodology

As seen in Fig. 2 (left), the sensor's data is a series of cloud points in 3D space over time. Essentially, the cloud points

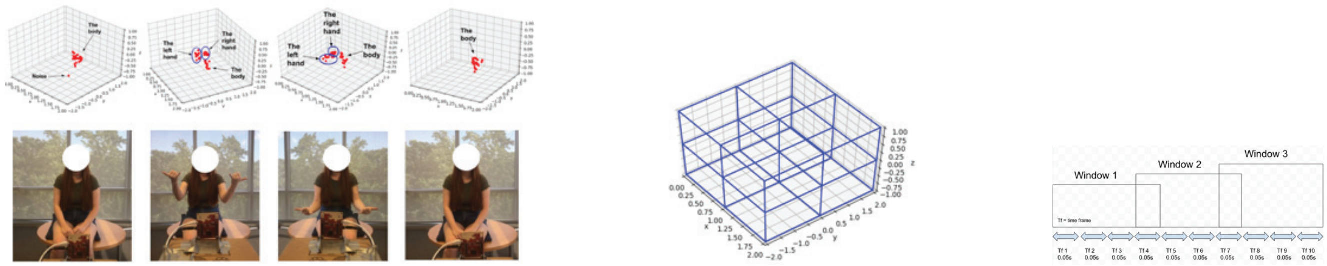

Figure 2: ”today”(left), Cell Division(mid) and Sliding Window(right) Visualization

were not tracking how the object moved. Instead, it illustrated the detection of the object in space. Therefore, continuous motion of the user cannot be determined by tracking the ”dancing” of the cloud points. However, almost all of the cloud points themselves carry information about the gestures. Therefore, in order to interpret this data, a large cube was constructed around the body, including all of the possible points that the user’s hands can reach in the 3D space. The range in each dimension of the cube was then equally divided into several parts (Fig. 2 mid). Next, the number of points in each cell were counted and normalized into the percentage of the total number of points in all the cells. A time frame for each second was not created. Instead, a certain number of seconds were combined in one time frame to make it a frame-by-frame series. In order to capture more connections of the gesture between the nearby windows, a sliding window was utilized which keeps some parts of the previous window in the following window, as shown in Fig. 2 right. Due to the noisy environment around the user, a density-based clustering algorithm was utilized to eliminate the noise which was too far from the body. After all of those processes, the cloud point time series sensor data was successfully transformed into a sample-feature dataset. Due to some cells of the cube not changing at all during the time frames, a variance threshold was applied to remove these features, ultimately increasing efficiency. Additionally, several reliable machine learning classification techniques were applied, such as K Nearest Neighbors, Support Vector Machine, and Random Forest to learn from the dataset.

Experiment Set-up & Results

As Table 1 showed, a list of commonly used signs were included in the proposed vocabulary. 30 samples of each gesture were collected per person from three professional ASL users with the same experiment set-up (20 cm from the sensor). Data was represented and trained with several reliable classification algorithms mentioned above. In short, the highest word prediction accuracy (95%) was achieved by utilizing Multilayer Perceptron (MLP). In terms of the cross-object prediction, the algorithm did not work without training on the different person’s data. It was found that the ”best” trade-off between different person’s data introduction and prediction accuracy was 30%, providing for an approximate prediction accuracy of 80%. However, the algorithm does have problems with recognizing similar gestures during cross-user prediction. Different users’ idiosyncrasies make the classifier harder to distinguish among similar gestures.

weather	alarm	set	call	today
reminder	order	on	off	yes
no	what	lock	movie	sports
help	hello	show		

Table 1: Vocabulary

Conclusion & Future Work

In this paper, a cell-division algorithm was proposed along with a sliding window technique to represent 3-D time series sensor data. It was successfully shown that this could be a solution for the problem of accurate, inexpensive, and noninvasive American Sign Language Recognition both in single-object and cross-object recognition. In the future, better preprocessing techniques can be developed, possessing certain markers which can help localize the hands among cloud points. Cells can also be adjusted based on the hands to make the cell-division algorithm more accurate. Furthermore, the gestures could be illustrated as a series of body-hand separated models in order to develop a way for the machine to better track the hands’ movement. Still, there is the question of if data needs to be interpreted in a human readable manner, or if the machine can learn from its binary world.

Acknowledgements

This project is sponsored by the George Mason University OSCAR Office. We would like to thank Frederick and Ari for ASL data collection as well as Dom, Panneer and anonymous reviewers for their valuable comments.

References

National Institute on Deafness and Other Communication Disorders (NIDCD). 2019. American sign language.

Potter, L. E.; Araullo, J.; and Carter, L. 2013. The leap motion controller: A view on sign language. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, OzCHI ’13, 175–178. New York, NY, USA: ACM.

Zafrulla, Z.; Brashear, H.; Starner, T.; Hamilton, H.; and Presti, P. 2011. American sign language recognition with the kinect. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI ’11, 279–286. New York, NY, USA: ACM.