Sign Language Recognition: Learning American Sign Language in a Virtual Environment

Jacob Schioppo
Zachary Meyer
Diego Fabiano
Shaun Canavan
University of South Florida
Tampa, FL, USA
jschioppo@mail.usf.edu
zacharymeyer@mail.usf.edu
dfabiano@mail.usf.edu
scanavan@usf.edu

ABSTRACT

In this paper, we propose an approach, for sign language recognition, that makes use of a virtual reality headset to create an immersive environment. We show how features from data acquired by the Leap Motion controller, using an egocentric view, can be used to automatically recognize a user signed gesture. The Leap features are used along with a random forest for real-time classification of the user's gesture. We further analyze which of these features are most important, in an egocentric view, for gesture recognition. To test the efficacy of our proposed approach, we test on the 26 letters

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3313025

of the alphabet in American Sign Language in a virtual environment with an application for learning sign language.

KEYWORDS

Gesture; sign language; virtual reality; classification

INTRODUCTION

Gesture recognition has been gaining increasing attention in the scientific community. This can be attributed to its applicability in fields such as robotics, human-computer interaction, sign language recognition, and interface design. It also has practical applications for real-time sign language recognition as both an interpretation device, as well as a learning tool. It has also seen success due to devices such as the Leap Motion sensor (Leap) [15].

One of the main applications that has seen success with the Leap is that of sign language recognition. The research has been carried out with applications in multiple languages such American sign language [5], [17], Arabic sign language [9], [14], and Indian sign language [11]. While much of the work using the Leap for sign language consisted of a single sensor, some work by Fok et al. [6] made use of two leap motions sensors. They fused the data collected by performing spatial alignment on both sensors and aligning them to the same reference coordinate system. Using this fusion method, they could classify 10 digits of the American sign language (0-9) with a recognition rate of 84.68%, compared to 68.78% with one sensor.

Gesture recognition has also shown success being incorporated with virtual reality (VR). Marchesi and Ricco have developed a wearable hand controller called GLOVR [12]. This controller has been designed with VR games in mind and makes use of a 9-axis Inertial Movement Unit to allow the user to interact with the virtual environment. While this type of device can allow for broad, smooth gestures, it does not include finger data as is required for many gestures such as those in sign language. Combining both VR and Vietnamese sign language recognition, Bich et al. [1] used a glove which includes ten flex sensors and one accelerometer to collect gesture data. From this gesture data they recognized a small subset of fourteen gestures. With this system, some of the gestures had a low recognition rate of approximately 64%, which is due largely to rotations around the z-axis.

Motivated by the success of gesture recognition in virtual reality, as well as the Leap Motion sensor [15], we propose an approach for sign language recognition that makes use of a virtual headset [8] and the Leap. The approach makes use of features from the Leap, captured from an egocentric view. We analyze which of these features are best, and test the efficacy of the proposed system on the 26 letters of the American Sign Language alphabet, in a virtual environment with an application for learning sign language.

LEAP MOTION FEATURES

The Leap uses two cameras and three infrared LEDs with a right-handed coordinate system that is measured in real world millimeters, where the Y-coordinate is pointing away from the camera. The Leap can infer the 3D position of a hand, making it a natural fit for sign language classification through feature descriptors. Studies have shown feature descriptors from the Leap using a desktop view [4], which we make use of for our proposed egocentric view. Namely the extended finger binary representation, max finger range, total finger area, and finger length-width ratio. We refer the reader to the work from Canavan et al. [4], for more details on these. Along with using these descriptors, we propose to use 3 complementary Leap features as detailed below.

Grab and Pinch Strength

The first feature, we propose to use, is the grab strength of the gesture, which shows how close the hand is to a fist in the range of [0, 1]. If the hand is open the grab strength will be 0.0, if the hand is closed the grab strength will be 1.0. The strength moves within this range as fingers curl to make a fist (Figure 1). Along with the grab strength, we also propose using pinch strength. A pinch is done between the thumb and another other finger of the hand. This is useful for letters such as *C* in the ASL alphabet. This feature is also in the range [0,1], with 0 being an open hand, and blends to 1, as a pinch between the thumb and any other finger occurs (Figure 1).

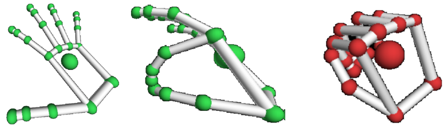

Figure 1: Pinch and grab. From left to right: open hand - pinch and grab strength are 0; pinch between thumb and index finger - strength is 1.0; grab - strength is 1.0

FingerTip Velocity

There are many cases in ASL where the gesture is dynamic such as the letters *J* and *Z*. The Leap provides access to the instantaneous 3D palm velocity in millimeters/second. To use this, we calculate the average velocity over the length of the dynamic gesture. When classifying the majority of ASL letters (24 of 26), the velocity will be close to 0 for all axes, however, for *J* and *Z* the velocity will be based on how fast the subject signs the letter. This velocity is useful for differentiating the ASL letter *I* from *J* and *D* from *Z* as the static features will be similar due to the shape of the letters (Figure 2).

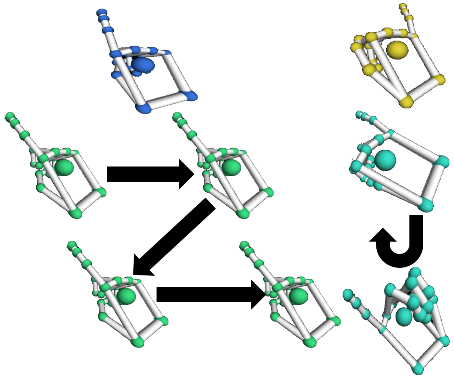

Figure 2: Leap data comparison of static and dynamic ASL letters. Left: *D* on top, with *Z* shown below. Right: *I* on top, with *J* shown below.

EVALUATION OF LEAP FEATURES FOR CLASSIFICATION OF SIGN LANGUAGE

To evaluate the features, we investigated a random forest and a deep feed forward network for sign language classification. For details on random forest, we refer the reader to the work from Brieman [3]. A brief overview and justification of our deep neural network is given below.

Deep Feedforward Neural Network

Recently, deep neural networks have gained increasing popularity for a variety of tasks including gesture [21] and posture [19] recognition, due in part to the work from Hinton et al. [7]. One variant of

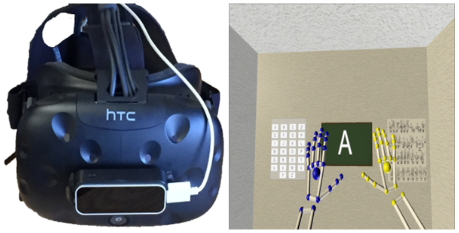

Figure 3: Left: Leap attached to HTC Vive; Right: view of subject while gesturing within FOV of LEAP.

deep networks, deep feedforward neural networks, have been shown to be powerful for hand gesture recognition in VR [22] abnormal breast detection in mammograms [20], and face verification [18].

For our experiments, we created a network with an initial input layer made of the same number of neurons as the size of our input vector, one hidden layer with number of neurons = ⌊(number of neurons in input layer + number of neurons in output layer)/2⌋, and finally an output layer with number of neurons equal to the number of classes (26 for our experiments). The softmax activation function was used, and the adamax optimizer [10] with a learning rate of 0.001.

Sign Language Classification from an Egocentric View

We evaluated the efficacy of the Leap features on the 26 letters of the ASL alphabet. To collect this data, we attached the Leap to an HTC Vive VR headset [8], with signer wearing the headset (Figure 3). The user is always able to see the Leap skeletal data as they move their hands within the Field of View (FOV) of the Leap (Figure 3). The Leap has a FOV of 150 degrees wide, while the HTC Vive has a FOV of 110 degrees, therefore the skeletal data (i.e. hands) will always appear within the virtual FOV.

We collected 780 instances, from 3 signers, of all 26 letters of the ASL alphabet (30 instances of each letter, 10 from each signer), and trained a random forest. Using 10-fold cross validation, we achieved a classification accuracy of 98.33% (767 of 780 instances correctly classified). The majority of the letters were classified with high accuracy, however, some letters such as *M* and *T* were incorrectly classified, which can be attributed to the similar nature of those letters. Our proposed velocity feature descriptor is also shown to have utility in recognizing dynamic gestures (e.g. *J* and *Z*), as *Z* was correctly classified 29 out of 30 times.

Using 10-fold cross validation with a deep feedforward network, We achieved an accuracy of 97.1% across the 26 letters of the ASL alphabet. It is interesting to note that while the deep network was 1.23% less accurate across all letters, it was able to better recognize the dynamic letters. The deep network achieved 100% accuracy for *J* and *Z*, while the random forest had an accuracy of 76.6% and 96.6% for *J* and *Z* respectively. This suggests that the dynamic Leap features (e.g. velocity) have more discriminative power for sign language classification when used with deep networks compared to random forests. While random forest and deep networks classify some letters differently, these results show the efficacy of the Leap features, from an egocentric view, for sign language classification. To further study these results, we calculated the correlation between each feature and the output class.

Correlation Between Features and Accuracy

While the results detailed are encouraging, it is important to know which of the Leap features are best for classifying ASL gestures. To investigate this, we calculated the correlation between each feature and the output classification. We did this by calculating the Pearson correlation coefficient

[16], which measures the linear correlation between two variables (e.g. feature and classification). By looking at this correlation we can rank each of the Leap features from highest to lowest.

The top 5 ranked features are (1) fingertip distance; (2) finger directions; (3) pinch strength; (4) grab strength; and (5) extended fingers binary representation. It is also interesting to note that velocity was ranked at the bottom, with the least amount of correlation, which makes sense as only 2 of the ASL letters had a velocity not close to 0, resulting in little information gain with the other 24 static ASL letters. The distance and direction features were found to be important as they give us information about the general shape of the hand, as well as the direction it is pointing, which is important when classifying and predicting ASL letters.

VIRTUAL SIGN LANGUAGE

Motivated by the power of the Leap features to classify ASL letters from an egocentric view, we propose an application to learn sign language in a virtual environment. VR-based instruction has been found to be an effective tool to improve the quality of learning outcomes [13]. We use the HTC Vive [8], along with the Leap [15] in an egocentric view (Figure 3), for our virtual learning environment.

To classify the ASL sign that the user is performing, we use the same experimental design as detailed in in the previous section. Again, due to the Leap having a FOV greater than the Vive, the user will experience an immersive experience, as well as accurate articulation of the sign. Once the user is in our virtual environment they are greeted with our virtual classroom. Our classroom contains a menu, where the user can select a letter of the ASL alphabet, a poster on the wall detailing how to correctly do each available gesture (e.g. 26 letters of the ASL alphabet), and a green chalkboard (Figure 4). In the virtual classroom, the user can select which letter of the alphabet they want to learn. by pointing to the corresponding button on the menu, which will be highlighted. The letter then appears on the chalkboard, to give immediate feedback that the learning process has begun

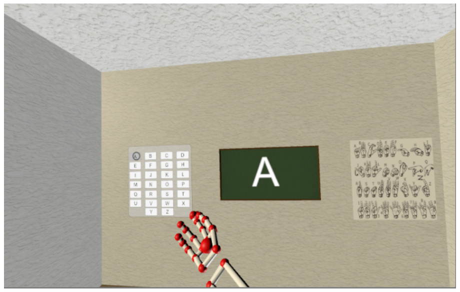

Figure 4: User selecting *A* in virtual classroom.

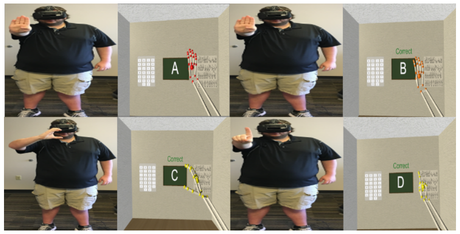

Figure 5: User learning sign language in our virtual environment. Top left: incorrect *A*. Top right: correct *B*. Bottom left: correct *C*. Bottom right: correct *D*.

Once a letter is selected the user has as much time as needed to perform the correct gesture. To encourage users to continue using our learning tool, we have taken a ”relaxed” approach to incorrect gestures. The user will only receive feedback for a correct gesture, vs. an incorrect one. This is done to provide positive feedback, and to mitigate any inaccuracies that may occur during classification. Due to the speed and reliability of random forests (with comparatively smaller training sets), we have chosen to use them for classification of the ASL letters, in VR. As can be seen in Figure 5, the Leap can accurately capture the gesture data (e.g. ASL letter) that is being signed resulting in an accurate tool for teaching sign language in a VR.

DISCUSSION AND FUTURE WORK

We proposed an approach to sign language recognition, along with a learning application in a natural and immersive virtual environment. While this application focused on one user in the virtual environment, we are interested in multiple people sharing the same virtual environment (i.e. a teacher and student). The proposed application can pave the way for remote teaching of sign language in VR. To evaluate transfer of knowledge, from the virtual environment, we will study the effectiveness of the user to generalize the learned sign language to the real-world [2].

We proposed using Leap features, from an egocentric view, to classify ASL (e.g. letters). A classification accuracy of 98.33% and 97.1% was achieved using a random forest, and a deep feedforward neural network, respectively. We also detailed an analysis of the correlation between the features and the output classification using the Pearson correlation coefficient. This allowed us to rank the features, showing which Leap features are most useful for an egocentric view.

REFERENCES

[1] D. Bich et al. 2016. Special Characters of Vietnamese Sign Language Recognition System Based on Virtual Reality Glove. In *Advances in Information and Communication Technology*. 572–581.

[2] C Bossard et al. 2008. Transfer of learning in virtual environments: a new challenge? *Virtual Reality* 12, 3 (2008), 151–161.

[3] L. Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[4] S. Canavan et al. 2017. Hand gesture rec. using a skeleton-based feature rep. with a random regression forest. (2017).

[5] C-H Chuan et al. 2014. American sign language recognition using leap motion sensor. In *ICMLA*. 541–544.

[6] K-Y. Fok et al. 2015. A real-time asl recognition system using leap motion sensors. In *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2015 International Conference on*. IEEE, 411–414.

[7] G. Hinton, S. Osindero, and Y-W. Teh. 2006. A fast learning algorithm for deep belief nets. *Neur. comp.* 18, 7, 1527–1554.

[8] HTC. 2018. *HTC Vive.* https://www.vive.com/us

[9] B. Khelil and H. Amiri. 2016. Hand gesture rec. using leap motion controller for rec. of Arabic sign language. (2016).

[10] D. Kinga and J. Adam. 2015. A method for stochastic optimization. *Intl. Conf. on Learning Representations* (2015).

[11] R. Mapari and G. Kharat. 2015. Real time human pose recognition using leap motion sensor. (2015).

[12] M. Marchesi and B. Riccò. 2016. GLOVR: a wearable hand controller for virtual reality applications. (2016).

[13] Z. Merchant et al. 2014. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education* 70 (2014), 29–40.

[14] M. Mohandes, M. Deriche, and J. Liu. 2014. Image-based and sensor-based approaches to Arabic sign language recognition. *IEEE transactions on human-machine systems* 44, 4 (2014), 551–557.

[15] Leap Motion. 2018. *Leap Motion.* https://www.leapmotion.com

[16] K. Pearson. 1895. Note on regression and inheritance in the case of two parents. *Proc. Royal Soc. of London* 58 (1895).

[17] L. Quesada, G. López, and L. Guerrero. 2015. Sign language recognition using leap motion. In *UCAml*.

[18] Y. Taigman et al. 2014. Deepface: closing the gap to human-level performance in face verification. *CVPR* (2014).

[19] A. Tang et al. 2015. A real-time hand posture rec. sys. using deep neural networks. *ACM Trans. on Int. Sys. and Tech.* 6, 2.

[20] S. Wang et al. 2017. Abnormal breast detection in mammogram images by feed-forward neural network trained by Jaya algorithm. *Fundamenta Informaticae* 151, 1-4 (2017), 191–211.

[21] D. Wu et al. 2016. Deep dynamic neural networks for multimodal gesture segmentation and recognition. *IEEE transactions on pattern analysis and machine intelligence* 38, 8 (2016), 1583–1597.

[22] D. Xu. 2006. A neural net. approach for hand gesture recognition in virtual reality driving training system of SPG. (2006).