**Gesture Interpreter**

Andrew Ladlow

Contents

1. [Introduction 3](#_Toc434278451)

[1.1 Motivation 3](#_Toc434278452)

[1.2 Aims 3](#_Toc434278453)

[1.3 Report Overview 4](#_Toc434278454)

1. [Background 5](#_Toc434278455)

[2.1 Leap Motion 5](#_Toc434278456)

[2.2 BSL in Education 5](#_Toc434278457)

[2.3 Gesture Recognition 5](#_Toc434278458)

[2.3.1 Dynamic Time Warping 5](#_Toc434278459)

[2.3.2 K-Nearest Neighbour & Support Vector Machines 6](#_Toc434278460)

[2.3.3 Hidden Markov Models 6](#_Toc434278461)

[2.3.4 Artificial Neural Networks 6](#_Toc434278462)

[2.3.5 $P Point-Cloud Recognizer 7](#_Toc434278463)

[2.4 Similar Applications 7](#_Toc434278464)

[2.4.1 LeapTrainer 7](#_Toc434278465)

[2.4.2 UNI 7](#_Toc434278466)

1. [Design 9](#_Toc434278467)

[3.1 A 9](#_Toc434278468)

1. [Implementation 10](#_Toc434278469)

[4.1 A 10](#_Toc434278470)

1. [The System in Operation 11](#_Toc434278471)

[5.1 A 11](#_Toc434278472)

1. [Testing and Evaluation 12](#_Toc434278473)

[6.1 A 12](#_Toc434278474)

1. [Conclusions 13](#_Toc434278475)

[7.1 A 13](#_Toc434278476)

[References 14](#_Toc434278477)

**Chapter 1**

# Introduction

## 1.1 Motivation

Human-computer interaction has, for many years, been limited to mouse pointing and typing based devices. With the more recent development of smartphones we have seen numerous breakthroughs with touch based devices. These methods are all inherently unnatural when compared with real world interaction, however. In all of these cases users must interact through some kind of interface as opposed to naturally communicating via hand signals or body language.

Currently, there are several released devices which aid in monitoring, recording and displaying these more natural forms of communication. Devices such as the ‘Microsoft Kinect’, ‘ASUS Xtion Pro’ and ‘Apple Primesense Carmine’ all feature some degree of motion capture – the issue lies in the accuracy of recognition methods, especially with more precise or intricate gestures. The gesture interpreter aims to utilise one of these devices, the Leap Motion, in order to facilitate natural gesture communication – particularly gestures found in the British Sign Language (BSL).

According to Rudak, et al [1] the Leap Motion is capable of recognising movement with an overall average accuracy of 0.7mm, a degree that devices such as the Kinect were unable to reach in comparison. In addition, a recent V2 firmware update released for the Leap Motion improved tracking performance and enhanced the features of its API.

The proposed application will aid users who wish to learn or reinforce their current knowledge of the BSL. The low cost of the Leap Motion device means it could be purchased by users themselves or incorporated into current sign language courses in order to help streamline the learning process.

## 1.2 Aims

The application should fulfil the following main objectives in order to achieve the goal described above:

* Record gestures performed by a user
* Recognise the gestures representing BSL alphabet characters and distinguish between them
* Display a matching gesture with a similarity score, given a gesture input

## 1.3 Report Overview

The structure of the report is as follows:

Chapter 1 discusses the problem and the proposed solution, including its aims. Chapter 2 analyses background research and related work as well as any similar applications. Chapter 3 … Chapter 4… Chapter 5… Chapter 6… Finally, chapter 7…

**Chapter 2**

# Background

## 2.1 Leap Motion

The use of the Leap Motion controller in order to detect sign language gestures was previously investigated by Araullo, et al [2]. The authors noted that although the controller showed potential, further development of the API was required – this was mainly due to inaccurate hand detection in certain scenarios e.g. pinching fingers together, interlocking hands or holding one hand above the other. Despite this, recognition improvement is still certainly possible; these tests were carried out prior to the release of the V2 firmware update for the Leap Motion in 2014 which led to improved tracking performance. As such it’s likely that the current accuracy of the device has also since improved.

## 2.2 BSL in Education

BSL courses such as those offered by the nationally accredited Signature [10] award a number of qualifications ranging from ‘Level 1’ to ‘Level 6’, with each level representing an incrementing complexity of known vocabulary required in order to qualify. Each level’s accompanying qualification is composed of a number of modules, each of which focus on a particular subtopic e.g. ‘BSL conversational skills’ and ‘Understand varied British Sign Language in a range of work and social situations’. All of these modules include guided contact time as well as additional work intended for private study. There is also an accredited online learning resource available for an additional cost.

The problem with this current system is that new users are incredibly reliant on their tutor for guidance – online resources help but only to a certain degree, users often need reassurance that they are performing gestures correctly and the instant feedback provided by the proposed software would do just that. In turn this would benefit both the students taking these courses and the schools offering them. The use of the low-cost Leap Motion could hence be used as an additional resource to augment the user’s understanding of the BSL in both tutor guided learning hours and private study.

## 2.3 Gesture Recognition

Probably the most important aspect of the system is its gesture recognition – as we’re dealing with data in real time, a suitable algorithm should calculate match results on the fly without causing delay or otherwise affecting the user when the application is running. Most approaches apply some form of machine learning on the Leap Motion frame data, or a specific subset of it. Generally it’s difficult to say with certainty that one algorithm is better than any other other due to the variance of testing conditions, input data etc. Some feasible examples are shown below.

#### 2.3.1 Dynamic Time Warping

The use of dynamic time warping (DTW) was explored by Russell, et al [3]. The appeal of DTW is that it isn’t reliant on the time taken or speed of each input in order to accurately compare them – this is particularly useful with gestures which are often performed at varying speeds. The authors demonstrate how DTW could be applied to 2D handwriting gestures and suggested that it could be extended for use with 3D data, i.e. gestures. They conclude that the DTW approach used is suitable for real-time comparison but it remains to be seen whether that is still the case when dealing with the increased complexity of gestures compared with just handwriting.

#### 2.3.2 K-Nearest Neighbour & Support Vector Machines

K-Nearest Neighbour (KNN) and Support Vector Machines (SVM) were proposed by Chuan, et al [7] as recognition algorithms for the American Sign Language using the Leap Motion. Tests were carried out using the 26 letters of the alphabet - results showed a recognition accuracy rate of 72.78% and 79.83% for the two methods, respectively. The authors mention some possible reasons for the low accuracy with both algorithms; compared with the BSL alphabet, the ASL is signed using only one hand – as a result some letter representations are very close to one another which led to misclassifications of the Leap Motion data. The use of BSL with these methods could show improved results as all of its gestures require two hands to perform and are therefore more varied.

#### 2.3.3 Hidden Markov Models

Markov models, in particular Hidden Markov Models (HMM), are generally known for their use in pattern matching algorithms for speech recognition or typing prediction. A Markov model is a network of states with each state being connected to another with a specific weight or probability. In a Markov model based system the future state of the system is only dependant on its current state and the probability of the states it’s linked to. A HMM differs in that its state is partially obscured. An example of this could be found in a speech recognition system where we are able to observe a waveform of speech but the actual spoken words is hidden. This can be compared to a gesture recognition system where we are given the movement data of a gesture but the actual intended gesture is hidden.

The use of a HMM was proposed by Chen [9] to support 2D and 3D motion recognition, achieving recognition rates of 91.9% in user-dependant testing and 96.9% in user- independent testing.

#### 2.3.4 Artificial Neural Networks

The use of artificial neural networks (ANN) was previously proposed by Mohandes, et al [8], in particular a Multilayer Perceptron neural network (MLP) for use with the Arabic Sign Language (ArSL). The proposed system resulted in a classification accuracy of over 99%. An artificial neural network is a type of machine learning algorithm which bears similarity to the human brain in that it is composed of a series of simple processing units, neurons. These neurons are interconnected and each of these connections have a determined weight. The network learns from experience when provided with test data, calculating specific outputs for given inputs.

It’s noted that the testing produced some erroneous results, similar to those found in the KNN/SVM with ASL testing described earlier. In this case this was due to fingers being occluded by the palm of the hand or by other fingers during recognition, as opposed to gestures just being too similar to one another to discern between. The authors suggest the use of a second Leap Motion device positioned to the side of the user. Combined with the Leap Motion in front of the user this should theoretically resolve the observed issues, though further work has yet to be carried out.

#### 2.3.5 $P Point-Cloud Recognizer

The $P recognizer ($P), designed by Wobbrock, et al [4], is another example of a gesture recognition algorithm. As described in the paper, the $P is “a 2-D gesture recognizer designed for rapid prototyping of gesture-based user interfaces”. The $P aims to overcome the complex task of matching user gestures by instead treating them as groups or “clouds” of points and evaluating each one in turn. Even the simplest of gestures could be created in many different ways depending on the properties of its strokes e.g. start and end points, order or time, and direction. The use of a point cloud helps remove any ambiguity from the gesture which simplifies comparison and recognition. According to the paper, the algorithm requires only 70 lines of code to function and delivered over 99% accuracy in user-dependant testing.

## 2.4 Similar Applications

#### 2.4.1 LeapTrainer

‘LeapTrainer.js’, created by O’Leary [5], is a browser based gesture and pose learning and recognition framework for the Leap Motion. Developed in JavaScript, LeapTrainer allows users to create and store gestures then replay them at will. The software recognises a gesture using a template matching algorithm, based off the $P recognizer discussed above.

The software supports both motion gestures and ‘pose’ gestures – the difference being that a motion gesture is the movement of one or both hands at or above a specific velocity, whereas a pose is the stationary position of one or both hands over a period of time. This distinct separation is particularly useful as it allows both of these of gestures to be recorded by the software, without the user having to explicitly define a type or be limited to one or the other.

In terms of improvement, LeapTrainer seems to be more of a proof of concept than a fully fleshed out application. The code could be adapted in a number of ways to suit varying requirements – for example, the gesture matchings could be transformed into speech output for a communication system.

#### 2.4.2 UNI

A commercial application, UNI, is currently in development and is scheduled to be released in summer 2016 by MotionSavvy [6]. UNI bares similarities to the proposed application in that it utilises the Leap Motion in order to translate gestures into spoken text. A proposed ‘crowd sign’ library would allow users to add and share gestures with other users via a cloud based dictionary system.

The software is closed source and based on a subscription model of $20 / month with an initial up front cost. This severely limits further adaptation or extension by like-minded developers, instead users are solely reliant on MotionSavvy to support the application in the future. It is currently unknown which operating systems UNI will support and how well the software will achieve the goals described on the website.

**Chapter 3**

# Design

## 3.1 A

Comparison of explicit and implicit gesture recognition / segmentation e.g. pressing keyboard to start stop vs. detection based on hand speed or position etc.

**Chapter 4**

# Implementation

## 4.1 A

Mention of frame drop / slowdown when comparing every bone direction between current frame and deserialized frame from file, similarly when just comparing palm positions and finger tip positions but to a lesser degree - not viable for real time. Must use a subset of relevant frame data as opposed to storing whole frames etc.

**Chapter 5**

# The System in Operation

## 5.1 A

A

**Chapter 6**

# Testing and Evaluation

## 6.1 A

A

**Chapter 7**

# Conclusions

## 7.1 A

Accuracy?

Limited to gestures not requiring body or face?

Must be stationary to use - possible combination with oculus VR or other head mounted display?

### References

[9] Chen, M. (2013). Universal motion-based control and motion recognition. On line publication, Georgia Institute of Technology, <http://www.dtic.mil/dtic/tr/fulltext/u2/a344219.pdf>. Last accessed 1 November 2015.

[7] Chuan, C. H., Regina, E., & Guardino, C. (2014, December). American Sign Language Recognition Using Leap Motion Sensor. In Machine Learning and Applications (ICMLA), 2014 13th International Conference on (pp. 541-544). IEEE, 2014.

[8] Mohandes, M., Aliyu, S., & Deriche, M. (2014, June). Arabic sign language recognition using the leap motion controller. In Industrial Electronics (ISIE), 2014 IEEE 23rd International Symposium on (pp. 960-965). IEEE. Chicago, 2014.

[6] MotionSavvy. (2015) UNI. Website, <http://www.motionsavvy.com/>. Last accessed 9 October 2015.

[5] O’Leary R. (2013) LeapTrainer.js. GitHub repository, <https://github.com/roboleary/LeapTrainer.js>. Last accessed 9 October 2015.

[2] Potter, L. E., Araullo, J., & 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 (pp. 175-178). ACM, 2013.

[10] Signature. (2015) British Sign Language. Website, <http://www.signature.org.uk/british-sign-language>. Last accessed 1 November 2015.

[4] Vatavu R.D., Anthony L., & Wobbrock J. O. (2012) Gestures as Point Clouds: A $P Recognizer for User Interface Prototypes. On line publication, University of Washington, <http://faculty.washington.edu/wobbrock/pubs/icmi-12.pdf>. Last accessed 13 October 2015.

[3] Vikram, S., Li, L., & Russell, S. (2013) Writing and sketching in the air, recognizing and controlling on the fly. In CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp. 1179-1184). ACM, 2013.

[1] Weichert, F., Bachmann, D., Rudak, B., & Fisseler, D. (2013) Analysis of the Accuracy and Robustness of the Leap Motion Controller. On line publication, PubMed Central, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3690061/pdf/sensors-13-06380.pdf>. Last accessed 16 October 2015.