**Gesture Interpreter**

Andrew Ladlow

Contents

1. [Introduction 3](#_Toc435486352)

[1.1 Motivation 3](#_Toc435486353)

[1.2 Aims 3](#_Toc435486354)

[1.3 Report Overview 4](#_Toc435486355)

1. [Background 5](#_Toc435486356)

[2.1 Leap Motion 5](#_Toc435486357)

[2.2 BSL in Education 5](#_Toc435486358)

[2.3 Gesture Recognition 6](#_Toc435486359)

[2.3.1 Dynamic Time Warping 6](#_Toc435486360)

[2.3.2 K-Nearest Neighbour & Support Vector Machines 6](#_Toc435486361)

[2.3.3 Hidden Markov Models 6](#_Toc435486362)

[2.3.4 Artificial Neural Networks 7](#_Toc435486363)

[2.3.5 $P Point-Cloud Recognizer 7](#_Toc435486364)

[2.4 Similar Applications 7](#_Toc435486365)

[2.4.1 LeapTrainer 7](#_Toc435486366)

[2.4.2 UNI 8](#_Toc435486367)

1. [Design 9](#_Toc435486368)

[3.1 User Requirements 9](#_Toc435486369)

[3.1.1 Use Case Diagram 9](#_Toc435486370)

[3.1.1 Use Case Descriptions 9](#_Toc435486371)

[3.2 Software Architecture Design 9](#_Toc435486372)

1. [Implementation 10](#_Toc435486373)

[4.1 A 10](#_Toc435486374)

1. [The System in Operation 11](#_Toc435486375)

[5.1 A 11](#_Toc435486376)

1. [Testing and Evaluation 12](#_Toc435486377)

[6.1 A 12](#_Toc435486378)

1. [Conclusions 13](#_Toc435486379)

[7.1 A 13](#_Toc435486380)

[References 14](#_Toc435486381)

**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 form of motion capture – the issue lies in their level of accuracy, especially with more precise or intricate gestures – e.g. movement of fingers as opposed to movement of arms. 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 Weichert, et al [12] 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 overall 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
* Given a gesture input, display a matching gesture with a similarity score

## 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 identifies the user requirements for the application, and subsequently details appropriate architectural designs. Chapter 4… Chapter 5… Chapter 6… Finally, chapter 7…

**Chapter 2**

# Background

## 2.1 Leap Motion

The Leap Motion is a palm sized USB device which tracks hands and fingers using optical sensors and infrared lights. The device was first released by Leap Motion, Inc. in July 2013[3]. A major software update, firmware version 2, was subsequently released for the device in May 2014[4]. The update claimed improved tracking performance and an enhanced API feature set. An image of the device is shown in figure 2.1.1.



**Figure 2.1.1:** Leap Motion device

The use of the Leap Motion controller in order to detect sign language gestures was previously investigated by Araullo, et al [8]. 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 – all of which created anomalous data due to the obscuration of finger positions. Despite this, recognition improvement is still certainly possible; these tests were carried out prior to the release of the aforementioned 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 [9] 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 [11]. 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 [2] 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 [1] 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 [5], 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 [10], 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 [7], 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 User Requirements

When considering the overall design of an application there are several sections which must be considered. The first of which is the clear definition of user requirements. In software development, a requirement is a property that a system must contain or exhibit in order for it to satisfy a user. Before any implementation occurs it’s crucial to ensure these requirements are clarified.

Requirements are grouped into two categories; functional and non-functional. The former describes the features of a system (what it does) whereas the latter describes how the system behaves (performance, reliability etc.). These requirements can be more easily identified via the creation of use case diagrams (shown in figure 3.1.1) and use case tables (figure 3.1.2). Diagrams aim to visualize the relationships between users of a system and possible use cases, as well as between use cases themselves. Use case tables specify the function of each use case, but don’t consider their implementation.



#### Figure 3.1.1: Use Case Diagram







#### Figure 3.1.2: Use Case Scenarios

From the use case analysis above the following functional and non-functional user requirements were established:

Functional

* R1.1: The system shall display a real time interpretation of the user’s hands during operation
* R1.2: A user shall be able to record and store their own image data for a given gesture
* R1.3: The system shall recognize a gesture provided by a user
* R1.4: The system shall present user feedback, a normalized score, based on the similarity between a given gesture and stored gestures

Non-Functional

* R1.5: The system’s real time display of a user’s hands shall be updated with a latency of 5ms or less
* R1.6: The recognition of gestures shall take no longer than 5ms to complete for each stored gesture
* R1.7: The system shall recognize gestures with an accuracy of at least 80%
* R1.8: The system shall implement a simplistic interface which is easy to use without relying on mouse and keyboard input
* R1.9 The system shall operate robustly (should not crash or otherwise close without user’s request)

## 3.2 Architectural Design

Considering the user requirements specified in the previous section, the application should consist of three main subsystems at the highest level:

* Leap Motion subsystem – Updates application with new data as the physical Leap Motion device generates it
* Gesture subsystem – Handles storage and recognition of performed gestures
* Visualization subsystem – Displays a visual interpretation of the current Leap Motion data, providing the user with feedback of their actions

This overall architecture to encompass all of these systems appropriately is shown in figure 3.2.1.

**Figure 3.2.1: Overall system architecture**

**Chapter 4**

# Implementation

## 4.1 Leap Motion Integration

The Leap Motion controller records tracking data in a series of frames. Each frame contains the positions of any detected hands or other pointable objects. Frames can be acquired by simply polling the device or via a call back method from an event listener which is assigned to the device. In the latter case, the Leap Motion will create a new thread for each new frame, and will pause execution until the current thread’s call back method (‘onFrame()’) has returned. This prevents an occurrence of thread-flooding, where threads are created at a faster rate than the device is able to process them.

Comparing the two integration choices, I found the second to be the most intuitive. The use of the event listener allowed me to manipulate or store the current frame’s data without having to consider the poll rate, which could in some cases be causing frames to be skipped, or duplicate frames to be requested, depending on the rate of frame polling compared with the rate of frame returns by the Leap Motion controller.

**4.2 GUI**

The application’s GUI is handled by JavaFX, a library of graphics and media packages which is written as a Java API, meaning it can be referenced from any standard Java based program. The properties of JavaFX allow external devices such as the Leap Motion to be easily incorporated within an application. Vos [13] described this combination in a schema shown in figure 4.2.1.



**Figure 4.2.1:** Interaction between Leap Motion and JavaFX

As described in Oracle’s documentation of JavaFX [14], a system using JavaFX will run two or more of the following threads at any given time:

* **JavaFX application thread**: Primary thread used by JavaFX. Essentially any content which can be seen by the user must be managed in this thread.
* **Prism render thread**: Allows the application to perform concurrent processing. While one frame is being rendered, the next can be pre-processed to help off-load the work required.
* **Media thread**: Runs in the background and synchronizes the latest frames using the application thread.

With JavaFX, any property that modifies a window’s live content can only be changed through a ‘Platform.runLater()’ call – this ensures that these modifications only occur on the application thread. The combination of this system and the Leap Motion listener lead to an approach where each call of ‘onFrame()’ includes one or multiple Platform.runLater() calls in order to modify the content shown in the JavaFX window. This ensures the application runs safely in regards to multithreading so as not to attempt to modify values in incorrect threads.

**4.2.1 Hand visualizer**

The inclusion of a visualization system to represent a user’s hands when they’re interacting with the application was a crucial initial design consideration. The user should be able to see their actions in real time, without having to swap between the Leap and the keyboard/mouse to e.g. navigate through menus – or having no visual feedback from the Leap entirely.

The visualization is setup through the use of a dedicated listener class (as described in section 4.1). When each new frame is captured by the Leap controller, this listener’s onFrame() method checks the content of the frame to find the number of hands (if any) captured in it. If there is at least one hand visible in the frame, the listener sets a Boolean flag which in turn has a listener associated with the GUI assigned to it. When this GUI listener is told there is content to display (when this flag is set)

**4.3 Gesture Recognition**

The application uses an adapted version of the $P recognizer (discussed briefly in section 2.3.5). The code was originally written in Javascript and C# for use with handwriting recognition but has been converted to the Java syntax and adapted to support points in three dimensions in order to function correctly with the Leap Motion.

At its highest level, the $P is an “instance-based nearest neighbour classifier with a Euclidean scoring function”. Breaking this down, the $P selects the most appropriate category for an object, given a selection of objects, by calculating the Euclidean difference in positions between all available objects in order to find the closest and hence the most likely match. The $P is instance-based, meaning it requires only a single training iteration in order to record data – this is in contrast to most other machine learning algorithms whose performance only improves when multiple training samples are provided.

In machine learning, objects are grouped into specific ‘categories’ based on a number of their ‘features’. Features refer to the properties of the object which make it unique, compared with other objects. In the context of the application, the features shown in table 4.3.1 are used to distinguish gestures from one another.

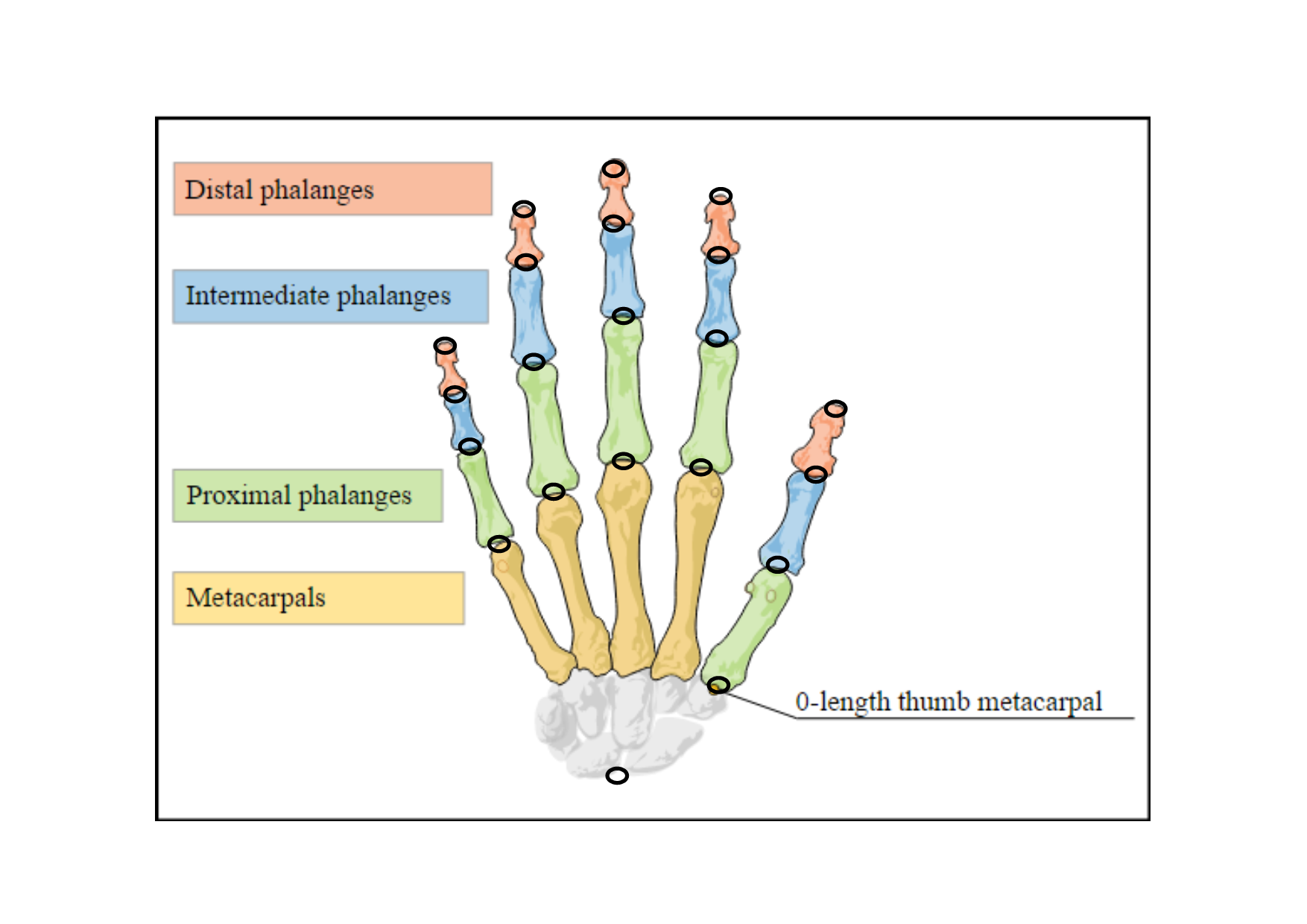
|  |  |  |
| --- | --- | --- |
|  | **Palm** | **Fingers** |
| **Features used** | Stabilized palm position | Stabilized tip position |
| Palm direction | Metacarpal end position |
| Palm normal | Proximal end position |
|  | Intermediate end position |
|  | Distal end position |

**Table 4.3.1:** Hand features used for gesture recognition

The meaning of each of these features is as follows:

* Stabilized palm position:
* Palm direction:
* Palm normal:
* Stabilized tip position:
* Metacarpal end position:
* Proximal end position:
* Intermediate end position:
* Distal end position:

Each position feature is highlighted in figure 4.3.2 from the LeapMotion API page [15].



**Figure 4.3.2:** Bone model with highlighted position features

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

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

[1] 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.

[2] 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.

[3] Etherington, D. (2013, April). Leap Motion Controller Ship Date Delayed Until July 22, Due To A Need For A Larger, Longer Beta Test. Website, <http://techcrunch.com/2013/04/25/leap-motion-controller-ship-date-delayed-until-july-22-due-to-a-need-for-a-larger-longer-beta-test/>. Last accessed 4 December 2015.

[4] LeapMotion. (2014) Leap Motion V2 Tracking Now in Public Developer Beta. Website, <http://blog.leapmotion.com/leap-motion-v2-tracking-now-in-public-developer-beta/>. Last accessed 16 November 2015.

[5] 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.

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

[8] 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.

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

[10] 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.

[11] 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.

[12] 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.

[13] Vos, J. (2014) Leap Motion and JavaFX. Website, <http://www.oracle.com/technetwork/articles/java/rich-client-leapmotion-2227139.html>.

Last accessed 2 February 2016.

[14] Oracle. (2014) JavaFX: Getting Started with JavaFX. Website, <https://docs.oracle.com/javase/8/javafx/get-started-tutorial/jfx-architecture.htm>.

Last accessed 2 February 2016.

[15] LeapMotion. (2015) Introducing The Skeletal Tracking Model. Website, <https://developer.leapmotion.com/documentation/java/devguide/Intro_Skeleton_API.html>.

Last accessed 5 February 2016.