**CHAPTER 1**

**INTRODUCTION**

This chapter contains a description of the hardware used and its application in different fields of computer vision. A present day scenario of how the Kinect hardware is used is described. The chapter then delves into the field of Sign Language Recognition and Computer Vision along with previous papers that have established pillars in these fields. The motivation behind this project along with the main and secondary objectives is clearly highlighted. The project timeline and updates is also mentioned.

**1.1 INTRODUCTION TO KINECT HARDWARE**

The Microsoft Kinect launched in November 2010, it was a targeted gaming controller for the Xbox 360 Gaming Console (Website, Official Microsoft). The device provides a controller free experience using computer vision techniques. The device is a combination of a RGB Camera, Depth Sensor and Microphone array. This caught attention in the developer community who wanted to repurpose the device to use it for specific computer vision projects as it was easily available and low cost at the time of release. Open source software was released unofficially followed by official support from Microsoft for Developers along with support libraries. It is able to track a user’s full body and provides framework for building sign language related applications.

The Microsoft Kinect has been used for a plethora of industry changing and out-of-the box gaming technology such as Kinect Sports, Kinect Fitness, Dance Central ,etc. that brought the convenience of all these facilities direct to your home. It has gone beyond just commercial applications as developers have adapted the device for various systems such as detecting human fall by measuring velocity, upper limb rehabilitation for people suffering from cerebral palsy, accessible design for the colour blind and games for children with cognitive disabilities.



Figure 1: A generic Kinect System

**1.2 PRESENT DAY SCENARIO IN SIGN LANGUAGE RECOGNITION AND COMPUTER VISION**

Sign language detection is generally done with one of two methods. These are either wearable assisted gesture recognition or computer vision methods. Computer Vision is a discipline that includes methods for acquiring, processing, analysing, and understanding images and, in general data to provide data or numerical information. Sign language wearables provide higher accuracy and better data to work with but are more of a hindrance than computer vision techniques that require bare minimum requirements from the user. In this decade, a lot of projects and research have been initiated with wearables. Sign Aloud, a wearable glove that leverages the power of motion detection embedded with a speech processor; works as a gesture to speech translator. Developed by students from University of Washington it is a system in development but with highly promising demos and test launches. Another system that was developed used electromyography to diagnose muscle movement and react accordingly with speech output corresponding to the sign gesture.

In the field of computer vision, a few novel solutions have been designed by various researchers. An array of camera combinations including Time of Flight cameras, RGB cameras, grayscale cameras have been used. Most of the research has been inclined toward ASL, BSL and sign language dialects from around the world. The degree of research in ISL has been very primitive (There are no open access ISL datasets available). A few papers have taken ISL into consideration and developed novel systems. A research study from Amrita School of Engineering has used feature extraction methods such as ALI and PCA to pioneer their work. There have been very few research studies that have combined insight into ISL with Artificial Intelligence.

**1.3 MOTIVATION**

Sign language is a gestural code representing the surrounding spoken language with varied features. These sign languages are complex natural languages and a part of the Deaf culture at all levels. Sign languages have various variants across the world. Our motive is to facilitate ISL words and capture them on Kinect sensors, thereby recognizing the dynamic ISL gestures. Previous research and projects have not delved into ISL along with effective use of Neural Networks/SVMs. With advancement in Machine Learning and computing capabilities, a dedicated system can be trained and tested for offline learning to identify gesture feature inputs.

The approach uses Kinect based global features for recognition making it a fine-grain description of a gesture. These features are statistically analysed and dimensionality reduction procedures are used. The features are then fed to a neural network that is later optimized to provide the most accurate classification. This system finds a variety of applications in:

* A brilliant application in the field of Computer Vision
* Widely used in Schools, Hospitals and Government agencies as a language interpreter
* An important tool in professional courses and associate programmes/degrees on Sign language.
* Sign language recognition systems are used in significant events like presidential speeches, courtroom affairs and office-work/conferences.

A positive end result would provide for a system that can classify a particular 3D gesture into a spoken phrase without human intervention or manual input at any stage.

**1.4 OBJECTIVE**

Primary objective of the research is to design a system that can classify 3 gestures captured using Kinect for Windows and display either text or speech outputs. The secondary objective would include gaining an insight into optimization procedures involved in Neural Networks and SVMs along with performance metrics used in these cases.

**1.5 TARGET SPECIFICATIONS**

Achievement of the primary as well as secondary objective would provide for a system that is stably built on the foundation of neural networks and large margin classifiers that can successfully classify a gesture into one of the three output classes. This system could find widespread use in schools where the hearing impaired generally require an interpreter for translation. It can find use in gatherings where the audience belongs to the hearing impaired community and needs to effectively communicate without third party intervention. It can be reverse engineered to learn sign language based on the training gestures provided.

**1.6 PROJECT SCHEDULE**

|  |  |
| --- | --- |
| *January 2015* | * Collection of training data * Feature extraction for obtaining Global Features * Variation of subsequent data sets and data accumulation * Logging of methods and procedures used |
| *February 2015* | * Algorithm implementation for obtaining final input data to feed to the network * Design of the neural network (EBPTA) for classification using binary data * Comparison of different algorithms with EBPTA for better classification results. |
| *March 2015* | * Experimentation with Neural Networks to identify best network for classification with the use of gesture co-ordinate data. * Variation of training data along with activation function to obtain best specificity and sensitivity results. |
| *April 2015* | * Finalization of neural network to be used and optimization processes. * Design of SVM to classify the gestures. |
| *May 2015* | * Optimization of SVMs * Documentation process. |

**1.7 ORGANIZATION OF THE REPORT**

Chapter 1 introduces the domain of sign language along with the hardware used (Microsoft Kinect). Chapter 2 describes the basics of image acquisition and relevant survey conducted in this field regarding the awareness and general statistics of physical features of the general population. It discusses the basic algorithm of pre-processing and neural networks. Chapter 3 describes the method incorporated for our project. It explains about binarization, Statistical analysis of gesture data incorporated with error back propagation and its different techniques. Chapter 4 deals with the result analysis by comparing the different techniques of classification. Chapter 5 is the conclusion of the result by comparing the different techniques of gesture classification.