CHAPTER 6

AUTOMATIC RECOGNITION OF STATIC AND DYNAMIC SIGN LANGUAGE GESTURES VIA KEY POINT DETECTION

Mok Xiu Yan and Zaid Omar

6.1 INTRODUCTION

Sign language is used as the primary tool for daily interactions by the deaf and hearing-impaired. Sign language can be defined as a structured form of hand gestures involving the application of different body parts including fingers, palm, body, arm, head, and facial expressions to convey ideas and information. However, due to its lack of usage, there remains an intrinsic difficulty faced by the hearing impaired to interact with the rest of society. In recent years, the development of robust tools and techniques for automatic sign language recognition has enabled us to potentially alleviate this communication gap. Though several challenges remain, namely that current systems mostly rely on external hardware such as Leap Motion (LM) and depth sensors which can be invasive and costly. In this chapter, our research is focused on developing a non-invasive sign language recognition system that can be adapted to desktop computing. The recognition uses the MediaPipe framework by extracting features using two-dimensional (2D) images and undergoing classification of hand movement, which can incorporate both static and dynamic hand gestures (Bazarevsky & Zhang, 2019; Hays & Mullen, 2020). The research scope is focused on the American Sign Language (ASL) as it represents the most ubiquitous and highly researched sign language in literature. The key point tracking step via MediaPipe can determine the precise three-dimensional (3D) key point localization of 21 hand-knuckle coordinates, 33 body pose key points and 468 face mesh key points inside the region of interest via regression, or direct coordinate prediction. The extracted features will then undergo classification using a neural network model. It is expected that the system can recognize static and dynamic gestures in real-time under robust lighting conditions, hand orientation, distance, complex background, and different users.

6.2 RELATED WORKS

Sign language is a structured form of hand gestures involving the application of different body parts including fingers, palm, body, arm, head, and facial expression to exchange ideas and (Sahoo et al., 2014). The first work on sign language recognition was identified to be in 1983, where Grimes of Bell Labs invented a digital data entry glove sewn with numerous bends, touch, and inertial sensors to obtain the motion of hands information (Cheok et al., 2019). It was designed specifically to recognize 26 manual gestures in the American Manual Alphabet. (Kramer & Leifer, 1990) subsequently created a communication aid called 'Talking Glove' that provides interaction for the deaf and hearing community. Similar designs have been developed over the years where some projects like CyberGlove and VPL DataGlove have gained attention in sign language translation. This was followed by the release of Microsoft Kinect and Leap Motion Sensors (LMS) which can provide 3D coordinates, contributing to a revolution in the recognition of 3D gestures. Despite these, however, glove-based sensors still share persistent flaws.

6.2.1 Feature Extraction

Widely used approaches for sign language feature extraction can be classified into sensor-based and vision-based techniques. In the work of (Almasre & Al-Nuaim, 2016), Kinect and LMS sensors are used to collect input data for a hand-gesturing model aimed to recognize 28 Arabic Sign Language gestures. LMS was again used by (Chuan et al., 2014) to apply Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN) algorithms in classifying 26 alphabets in ASL. The same sensor was used with two digital cameras by ElBadawy (2015) to collect data. The generation of text output was done with 20 hand gestures of different words and fed into the feature extraction and sign language interpreting algorithm for the recognition of text data.

The CyberGlove device consists of proprietary resistive bend-sensing technology to transform hand and finger motion into 22 points of high accuracy joint angle data in real-time. The glove has 22 sensors, four abduction sensors, and three flexion sensors per finger, palm arch sensor with another sensor to calibrate abduction and flexion. In the experiment by Kong and Ranganath (2018), the gesture data was obtained by the magnetic trackers, and CyberGlove successfully categorized 28 sign gestures using Vector Quantization Principal Component Analysis (VQPCA). Another microelectromechanical sensor glove named ADXL 202 is a low-power and economical device that comes with two-axis accelerometers on an integrated circuit. The ADXL 202 accelerometers are fitted onto a glove, one at the back of the palm and five at each finger.

Conversely, fully vision-based approaches in sign language recognition applications are relatively rare. A study carried out by Cheok et al. (2020) developed a sign language recognition app running on an iOS smartphone. The segmentation of the hand region from the background was done using Canny edge detection and skin color segmentation with histogram back-projection. Convexity effect, K-curvature and Hough line techniques were applied to obtain the shape feature of hands, while the chain code method was implemented to