Engineering Applications of Artificial Intelligence 24 (2011) 1204–1213

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

American Sign Language word recognition with a sensory glove using artificial neural networks

Cemil Oz [a,*], Ming C. Leu [b]

[a] Department of Computer Engineering, Computer and Informatics Faculty, University of Sakarya, Sakarya, Turkey
[b] Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, Missouri, USA

ARTICLE INFO

Article history:
Received 10 February 2011
Received in revised form
29 April 2011
Accepted 29 June 2011
Available online 31 July 2011

Keywords:
American Sign Language (ASL)
ASL recognition
Hand-shape recognition
Finger spelling recognition
Artificial Neural Networks (ANNs)

ABSTRACT

An American Sign Language (ASL) recognition system is being developed using artificial neural networks (ANNs) to translate ASL words into English. The system uses a sensory glove called the Cyberglove™ and a Flock of Birds® 3-D motion tracker to extract the gesture features. The data regarding finger joint angles obtained from strain gauges in the sensory glove define the hand shape, while the data from the tracker describe the trajectory of hand movements. The data from these devices are processed by a velocity network with noise reduction and feature extraction and by a word recognition network. Some global and local features are extracted for each ASL word. A neural network is used as a classifier of this feature vector. Our goal is to continuously recognize ASL signs using these devices in real time. We trained and tested the ANN model for 50 ASL words with a different number of samples for every word. The test results show that our feature vector extraction method and neural networks can be used successfully for isolated word recognition. This system is flexible and open for future extension.

1. Introduction

Sign language, which is a highly visual–spatial, linguistically complete and natural language, is the main mode of communication among deaf people. However, deaf people still experience serious problems communicating with people who hear normally, almost all of whom do not understand sign language systems such as American Sign Language (ASL). This communication barrier affects deaf people's lives and relationships negatively. Deaf people usually communicate with hearing people either through interpreters or text writing. Although interpreters can facilitate communication between deaf persons and hearing persons, they are often expensive, and their involvement leads to a loss of independence and privacy. While writing is used by many deaf people to communicate with hearing people, it is very inconvenient while walking, standing at a distance, or when more than two people are involved in a conversation.

Sign language is not universal. Different countries have different sign languages, for example, American Sign Language (ASL) and German Sign Language (GSL) have different alphabets and word sets. The similarities among signs in a sign language are created by complex body movements, i.e., using the right hand, the left hand, or both. When signs are created using both hands, the right hand is more active than the left hand. Sign language speakers also support their signs with their heads, eyes, and facial expressions.

Many researchers have been working on the recognition of various sign languages and gestures, but this research poses major difficulties due to the complexity of hand and body movements in sign language expression. Sign language recognition research can be categorized into three major classes: (i) computer-vision based, (ii) data-glove and motion-sensor based, and (iii) a combination of these two methods. Computer-vision based ASL recognition relies on image processing and feature extraction techniques for capturing and classifying body movements and handshapes when a deaf person makes an ASL sign. On the other hand, data-glove and motion-tracker based ASL recognition methods use a sensory glove and a motion tracker for detecting handshapes and body movements. The third method includes a combination of techniques from these two methods (Oz et al., 2004).

Acquiring data is more difficult with the vision-based method than with the data-glove and motion-sensor based methods. Data can be collected efficiently through a 3-D vision system, which has multiple cameras and a fast frame grabber. This system requires complicated image processing methods, which demand more data and slow the recognition rate. The main advantage of this approach is that the user does not need to wear any uncomfortable devices. Additionally, facial expressions can be incorporated. In the data-glove and motion-sensor based systems,

* Corresponding author. Tel.: +90 264 295 5598; fax: +90 264 295 5601.
E-mail addresses: coz@sakarya.edu.tr (C. Oz), mleu@mst.edu (M.C. Leu).

0952-1976/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2011.06.015

the signer has to wear a glove and sensor devices that measure the physical features of the gesture, e.g., trajectory, angles, motion, and finger bending.

In the earliest linguistic description of ASL, Wilbur (1987) used a structural linguistic framework to analyze sign formation. His purpose was to develop a national system for writing signs that contained symbols for each individual hand shape, location, and movement. Stokoe (1978) analyzed ASL formation and suggested additions to the three basic building blocks of hand shape, location, and movement. One major parameter, orientation of the palm, was suggested by Battison (1978). These and similar ASL studies have furthered the research of other ASL recognition scientists.

In parallel to the advancements in sensor and computer technology, some successful computer-vision based sign language recognition systems have been developed. Earlier sign language recognition research appeared in the literature at the beginning of the 1990s. Charayaphan and Marble (1992) developed an image-based processing system to understand ASL using hand motions. Takahashi and Kishino (1991) used a range classifier to recognize 46 Japanese Kana manual alphabet with a VPL data glove. Their study was based on simply encoding data ranges for joint angles and hand orientation.

Since 1990, Artificial Neural Networks have been used widely for solving engineering and industrial problems. Because of the popularity of ANNs, sign language researchers have applied this algorithm to solve their problems. Kramer and Leifer developed an ASL finger spelling system using a Cyberglove with the use of a neural network for feature classification and sign recognition (Kramer and Leifer, 1990; Kramer, 1996). Murakami and Taguchi (1991) established a recurrent neural network method to recognize 110 distinct Japanese Sign Language signs. Waldron and Kim (1995) used a neural network method to recognize 14 ASL signs using a different network for hand shape and hand orientation and position. Wysoski et al., (2002) developed an image-based ASL recognition system with a neural network for 26 static postures. Allen et al. (2003) developed an ASL finger spelling recognition system, which could recognize 24 ASL letters with a neural network. Wang et al. (2004) designed an ASL gesture recognition system with a sensory glove using an ANN, a Hidden Markov Model (HMM), and a minimum distance classifier. Oz and Leu (2007) designed an ASL recognition system based on linguistic properties with a sensory glove using an ANN.

The Hidden Markov Model, which has a well-founded mathematic basis and is an efficient doubly stochastic process, has been used widely in speech recognition, text recognition, and other engineering problem solving (Rabiner, 1989; Takiguchi et al., 2001). Many ASL researchers achieved successful results using HMM. Vogler and Metaxas (1997, 1999) used HMM for continuous ASL recognition with video streaming. In 1997, they were able to recognize 53 signs and a completely unconstrained sentence structure. In 1999, they were able to recognize ASL sentences with 22 signs based on ASL phonemes. Grobel and Assan (1996) used HMMs to recognize isolated signs based on computer vision with the signers wearing colored, normal gloves. Their accuracy was 91.3% for 262 signs.

There are also some Human Computer Interaction (HCI) studies based on human gestures. Lee and Xu (1996) used HMM to recognize the ASL alphabet for a human–robot interface. Lee et al. (2000) developed a hand gesture recognition system with human–computer interaction. Stergiopolou and Papamarkos (2009) used a neural network with a shape fitting filter to recognize hand gestures.

In this study, we present an ASL word recognition system that is constructed to translate ASL signs into the corresponding English words with an ANN method. A reliable adaptive filtering system with a recurrent neural network is used to determine the duration of ASL signing. All parameters affect the accuracy of feature vector extraction. The histogram method is used to extract features from ASL signs. Based on these features, a word recognition neural network is used as a classifier to convert ASL signs into English words. The developed system is capable of recognizing all 50 ASL words used in the testing.

2. System hardware and software

One of the primary means by which we physically connect to the world is through our hands. We perform most of our everyday tasks with them; however, along with our hands, we also rely on devices such as a mouse, keyboard, and joystick to work with computers and computer applications. Glove-based input devices could overcome this limitation (Sturman and Zelter, 1994). Commercial devices such as the VPL data glove and the Mattel power glove have led to an explosion of research and development projects using electronic glove interfaces for computer applications with computer controlled devices. Examples of these applications include virtual reality, video games, scientific visualization, puppetry, and gesture-based control.

We use a right-hand Cyberglove^{TM} (Fig. 1) to retrieve the joint angles for gesture features. The glove has 18 sensors, which measure the bending angles of fingers at various positions. We use 15 sensors on the glove: three sensors for the thumb, two sensors for each of the other four fingers, and four sensors between the fingers. The frequency of data collection is up to 150 Hz.

To track the position and orientation of the hand in 3-D space, the Flock of Birds® motion tracker (Fig. 2) mounted on the hand and wrist is used. The receiver is located in a DC pulsed magnetic field, and the effective range is up to 8 ft around the transmitter. The measuring frequency is up to 144 Hz.

Open Inventor SDK (Software Development Kit) is used in the software development for the 3-D scene rendering. It is a high-level toolkit developed in OpenGL for graphic rendering and user interaction. We use the Microsoft® Speech SDK for the programming of speech synthesis. The software system is implemented using Object Oriented Programming (OOP) technology; therefore, it is easily extendable.

Fig. 3 shows the overall structure of our system. The Cyberglove sensory glove and the Flock of Birds motion tracker are connected to the computer system with two separate RS-232 serial ports. The data stream from these devices is retrieved and

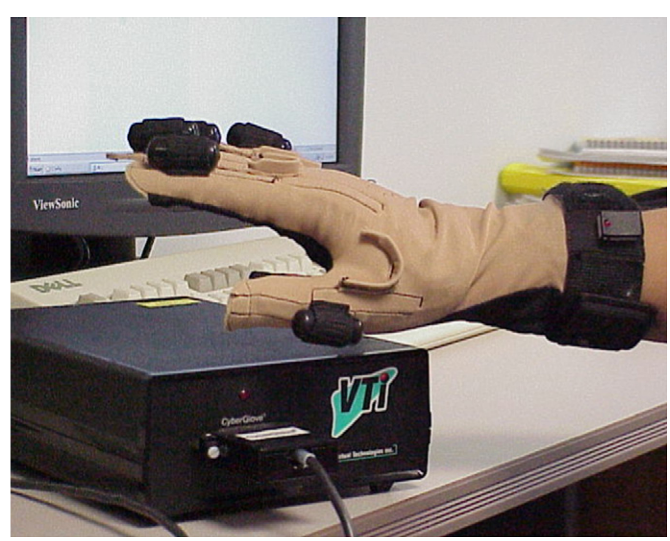

Fig. 1. Cyberglove^{TM} with 18 sensors.