International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 2, February 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training: Survey

Jeni Moni[1], Anju J Prakash[2]

[1]M. Tech Student, [2]Assistant Professor,

[1,2]Department of Computer Science and Engineering, Sree Buddha College of Engineering, Kerala, India

ABSTRACT

Sign Language (SL) is a medium of communication for physically disabled people. It is a gesture based language for communication of dumb and deaf people. These people communicate by using different actions of hands, where each different action means something. Sign language is the only way of conversation for deaf and dumb people. It is very difficult to understand this language for the common people. Hence sign language recognition has become an important task. There is a necessity for a translator to communicate with the world. Real time translator for sign language provides a medium to communicate with others. Previous methods employs sensor gloves, hat mounted cameras, armband etc. which has wearing difficulties and have noisy behaviour. To alleviate this problem, a real time gesture recognition system using Deep Learning (DL) is proposed. It enables to achieve improvements on the gesture recognition performance.

KEYWORDS: *Sign Language, Deep Learning (DL), Gesture Recognition*

How to cite this paper: Jeni Moni | Anju J Prakash "A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training: Survey" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-2, February 2020, pp.620-623, URL: www.ijtsrd.com/papers/ijtsrd30032.pdf

IJTSRD30032

INRODUCTION

The sign language is a very important way for communication with deaf and dumb people. Each gesture has a specific meaning in sign language. SL is considered basically as a non-verbal language Lots of research is going on image based approaches only because of advantage of not need to wear complex devices like Hand Gloves, Helmet etc. Sign recognition is related as image understanding. Sign detection and sign recognition are the two major phases. Sign detection can be defined as the process of extracting feature of certain object with respect to certain parameters. Sign recognition is the process of recognizing a certain shape that differentiates the object from the remaining shapes.

Sign language is often considered as the most grammatically structured gestural communications. This nature makes SL recognition an ideal research field for developing methods to address different problems such as human motion analysis, human-computer interaction (HCI).The problem of communication has been addressed by several companies and researchers, who have provided their solution. However, the problem is still under considerable attention.

Figure1. Sign Language

Previous methods employs sensor gloves, armband etc. which has wearing difficulties and have noisy behaviour. To solve this problem, a real time gesture recognition system using deep learning is proposed. Therefore it is possible to achieve improvements on the recognition performance. This will helps common people for recognizing gestures and to communicate with deaf or dumb people.

LITERATURE SURVEY

Various methods are used for sign language recognition. Some of them are discussed below.

1. Sign Language Recognition Using Sensor Gloves

S. A. Mehdi et al. [2] proposed a sign language recognition approach using sensor gloves. Sensor gloves are normally gloves made out of cloth with sensors fitted on it. Using data glove is a better idea over camera as the user has flexibility of moving around freely within a radius limited by the length of wire connecting the glove to the computer, unlike the camera where the user has to stay in position before the camera. This limit can be further lowered by using a wireless camera. The effect of light, electric or magnetic fields or any other disturbance does not affect the performance of the glove. 7-sensor glove of 5DT Company is used. It has 7 sensors on it. 5 sensors are for each finger and thumb. One sensor is to measure the tilt of the hand and one sensor for the rotation of the hand. Optic fibers are mounted on the glove to measure the flexure of fingers and thumb. Each sensor returns an integer value between 0 and 4095. This value tells about the bent of the sensor. 0 means fully stretched and 4095 means fully bent. So, a range of 7 * 4096 combinations obtained as input.

Artificial Neural Network with feed forward and back propagation algorithms have been used. Feed forward algorithm is used to calculate the output for a specific input pattern. Three layers of nodes have been used in the network. First layer is the input layer that takes 7 sensor values from the sensors on the glove. So this layer has 7 sensors. Next layer is the hidden layer, which takes the values from the input layer and applies the weights on them. This layer has 52 nodes. This layer passes its output to the third layer. The third layer is the output layer, which takes its input from the hidden layer and applies weights to them. There are 26 nodes in this layer. Each node denotes one alphabet of the sign language subset. This layer passes out the final output. A threshold is applied to the final output. Only the values above this threshold and considered.

2. A Wearable Hand Gloves Gesture Detection Based On Flex Sensors for Disabled People

Purohit et al. [3] introduced a wearable hand gloves gesture detection based on flex sensors for disabled people. The data glove is fitted with flex sensors along the length of each finger. The flex sensors output a stream of data that varies with degree of bend. The analog outputs from the sensors are then fed to microcontroller. It processes the signals and perform analog to digital signal conversion. The gesture is recognized and the corresponding text information is identified. The user need to know the signs of particular alphabets and he need to stay with the sign for two seconds. There are no limitations for signs it is hard to build a standard library of signs. The new sign introduced should be supported by the software used in the system. These sensors are attached along the fingers. The degree of bending of fingers and thumb produces the output voltage variation which in turn on converting to analog form produces required voice. A pair of gloves along with sensors enables mute people or old people to interact with the public in the required sentence which is very much helpful for them. At server side the system takes the input from the micro controller and based on the combination of those inputs it will match the pattern with already fed pattern in the database and if the pattern is not available in database, the system will respond with “not available” value.

3. Sign Language Recognition Using Image Based Hand Gesture Recognition Technique

A. S. Nikam et al. [4] proposed image based hand gesture recognition technique. Image based gesture recognition system is divided into three steps In Image-preprocessing color to binary conversion & Noise filtering is done for captured image. The set of operations which performs on the image based on shapes are known as Morphological operations. There are two most basic morphological operations: Erosion and Dilation, it uses for Removing noise, Separation of individual elements and joining misaligned elements in an image, even Finding of intensity bumps or holes in an image. Erosion shrinks boundaries of an image and enlarges holes. Erosion can be used to remove noises from an image. And Dilation is used to add pixels at region of boundaries or to fill in holes which generate during erosion process. Dilation can also be used to connect disjoint pixels and add pixels at edges. Tracking is mainly used for tracking a hand gesture from capture image using Convexity hull algorithm. Finally recognition is done with the help of features like convex hull and convex defects taken from tracking.

4. Real-Time Hand Gesture Recognition with EMG Using Machine Learning

A. G. Jaramillo et al. [5] proposed hand gesture recognition with EMG using machine learning. Myo armband is a sensor which is used because of the following reasons: low cost, small size and weight, software development kit (SDK) and because the Myo is a small and open source sensor that is easy to wear. The Myo armband has eight EMG surface dry sensors, and an inertial measurement unit (IMU). The eight surface sensors measure 200 samples per second of the electrical activity of the muscles. The IMU has 9 degrees of freedom (accelerometer, gyroscope, and orientation in the X, Y, and Z-axes). The Myo armband uses Bluetooth technology for transmitting the data to the computer. Finally, the Myo armband has incorporated a proprietary system capable of recognizing five gestures of the hand: pinch, fist, open, wave in, and wave out. EMG is a measure of the electrical activity produced by the muscles of the human body. The EMG signal is a linear summation between several trains of MUAPs. The amplitude and frequency of the EMG signals are affected by on the muscular fatigue, the age of the person, neuromuscular diseases, and the temperature and the thickness of the skin.

For feature extraction, different techniques in time, frequency, and time-frequency domains to obtain meaningful information are applied. In the time domain, features like the mean absolute value, nth-order autoregressive coefficients, zero crossing, length of the signal, sign of slope changes, modified mean absolute value, simple square integral are tested. In the frequency domain, features like the power spectrum, mean and median frequencies, frequency histogram, mean power, and spectral moments are tested. In the time-frequency domain, feature like the wavelet transform is tested. The classification stage determines to which class (gesture) a feature vector extracted from the EMG signals belongs to. The most common classifiers used in the hand gesture recognition with EMG are support vector machines and neural networks.

5. Real Time Indian Sign Language Recognition System to Aid Deaf-Dumb People

P. S. Rajam et al. [6] proposed a real-time sign language recognition system to aid deaf and dumb people. The

proposed method uses 32 combinations of binary number sign are developed by using right hand palm image, which are loaded at runtime. An image captured at run time is scanned to identify fingertip positions of the five fingers namely little fingers, ring finger, middle, index finger and thumb finger. The tip of fingers is identified by measuring their heights with respect to a reference point at the bottom of the palm close to the wrist. The heights are determined by Euclidean distance measurements. Number of instances in a scan are less than or equal to 3 in the case of Left-Right Scan and it is less than or equal to 2 in the case of Right- Left Scan which are determined by the UP or DOWN positions of the fingers. The output is obtained originally in the form of binary string of length of five in which the most significant bit represents the LITTLE finger and the least significant bit represents the THUMB finger. The string is then coded into the equivalent decimal numbers.

6. Sign Language Recognition Using Principal Component Analysis

A. Saxena et al. [7] proposed sign language recognition using principle component analysis. Principal component analysis is a fast and efficient technique for recognition of sign gestures from video stream. It is a rather general statistical technique that can be used to reduce the dimensionality of the feature space. Capturing of images from live video can be done using webcam or an android device. In this proposed technique, it is possible to capture 3 frames per second from video stream. After that we compare three continuous frames to know the frame containing static posture shown by hand. This static posture is recognized as a sign gesture. Now it is matched with stored gesture database to know its meaning. This system has been tested and developed successfully in a real time environment.

7. Using Multiple Sensors for Mobile Sign Language Recognition

H. Brashear et al. [8] proposed sign language recognition using multiple sensors. Multiple sensor types are used for disambiguation of noise in gesture recognition. In this case, accelerometers with the three degrees of freedom, mounted on the wrists and torso to increase our sensing information are used. The accelerometers will capture information that the vision system will have difficulty with such as rotation (when hand shape looks similar) and vertical movement towards or away from the camera. The camera will provide information not gathered by the accelerometers such as hand shape and position. Both sensors collect information about the movement of the hands through space. By adding multiple sensor types, the accuracy of the system will be improved in noisy or problematic conditions. It is important to add that sensor selection is based on the amount of information the sensor collects and its "wear- ability".

The current system could be partially concealed by embedding the camera in a normal hat, such as a baseball cap, and combining visual markers and accelerometers into a watch or bracelet. Proposed system consists of a wearable computer, heads–up display, hat–mounted camera, and accelerometers. The system captures video of the user signing along with accelerometer data from the wrists and body. The left hand is marked by a cyan band on the wrist and the right hand is marked by a yellow band. The HTK component of the system has been redesigned using the Georgia Tech Gesture Toolkit, which provides a publicly available toolkit for developing gesture–based recognition systems.

8. Real-Time Sign Language Recognition Based On Neural Network Architecture

P. Mekala et al. [9] proposed real- time sign language recognition based on neural network architecture. The video sequence of the signer, i.e. the person conveying in the sign language, can be obtained by using a camera. The initiation of the acquisition is being done manually. Local changes due to noise and digitization errors should not radically alter the image scene and information. In order to satisfy the memory requirements and the environmental scene conditions, preprocessing of the raw video content is highly important. Under different scene conditions, the performance of different feature detectors will be significantly different. The nature of the background, existence of other objects (occlusion), and illumination must be considered to determine what kind of features can be efficiently and reliably detected. Usually the hand shape and the movement are of major concern in order to guess the word/sentence. The feature vector is a single row column matrix of N elements. The feature vector computation involves time and memory.

Training and generalizing are the most basic and important properties of the neural networks. The neural network architecture consists of three layers - an input layer, one hidden layer and an output layer. In the gesture classification stage, a simple neural network model is developed for the recognition of gestures signs using the features computed from the video captured. The features can then be extracted from the video captured using any of the following system. Sign language recognition using neural networks is based on the learning of the gestures using a database set of signs. There is necessity of universal database as the applications grow and in that case such sequential search algorithms fail to meet the timing and memory constraints.

9. Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique

Joyeeta et al. [10] proposed sign language sign language recognition using Eigen value weighted Euclidean distance based classification technique. Eigen values and Eigen vectors are a part of linear transformations. Eigen vectors are the directions along which the linear transformation acts by stretching, compressing or flipping and Eigen values gives the factor by which the compression or stretching occurs. In case of analysis of data, the Eigen vectors of the covariance are being found out. Eigenvectors are set of basis function which describes variability of data. And Eigen vectors are also a kind of coordinate system for which the covariance matrix becomes diagonal for which the new coordinate system is uncorrelated. The more the Eigen vectors the better the information obtained from the linear transformation. Eigen values measures the variance of data of new coordinate system. For compression of the data only few significant Eigen values are being selected which reduces the dimension of the data allowing the data to get compressed The first phase for proposed system is the skin filtering of the input image which extracts out the skin colored pixels from the non-skin colored pixels. The input RGB image is first converted to the HSV image. The motive of performing this step is RGB image is very sensitive to change in illumination

condition. The HSV color space separates three components: Hue which means the set of pure colors within a color space, Saturation describing the grade of purity of a color image and Value giving relative lightness or darkness of a color. Next phase is the cropping of hand. For recognition of different gestures, only hand portion till wrist is required, thus the unnecessary part is clipped off using this hand cropping technique. Classifier was needed in order to recognize various hand gestures.

10. Vision-Based Hand Gesture Recognition System for a Dynamic and Complicated Environment

C. Liao et al. [11] proposed Vision-Based Hand Gesture Recognition System for a Dynamic and Complicated Environment. The proposed system consists of four stages, detection of the appearance of hands, segmentation of hand regions, detection of full palm, and hand gesture recognition. Detection of the appearance of hands is to find out when a hand appears in the front of the camera. Moreover, some morphological techniques, along with two stage skin color detection, are employed to alleviate the effect of noise. The proposed two-stage skin color detection approach is adopted from the idea of handling outliers to extract the palm from a complicated background. Following that, detection of full palm is conducted to know whether the hand reaches beyond the field of the camera view. The concept of ergonomics is employed to determine whether the hand is beyond the field of the camera view.

In vision-based human-computer interface systems, the segmentation of foreground objects such as hands, faces, and so on from the background is a major issue. Skin color detection technique is to extract foreground objects from the background image based on color information. This method segments foreground objects only by their color information without considering their shapes. The user may wear wrist artifacts or rings, and skin-color-like noises are allowed to exist in the background as long as they are smaller than the hand. Hence, background noises often exist after the skin color pixels of input images are detected. After the palm region is segmented from the background and filtered by the algorithm proposed, the palm is obtained and then the system needs to know how many fingers have been raised. In order to detect the number of fingers, the hand of the binary image will be transformed into a polar image for recognition.

CONCLUSION

The problem of communication has persisted for the people who are physically disabled such hearing impairment or physically mutes. The problem has been addressed by various researchers. However, the problem is still under considerable attention because of its feasibility and availability. Previous methods employ sensor gloves, armband, helmets etc. which has wearing difficulties and have noisy behaviour. To alleviate this problem, a real time gesture recognition system using deep learning is proposed. This enables to achieve improvements on the recognition performance.

REFERENCES

[1] R. Cui, H. Liu and C. Zhang, "A Deep Neural Framework for Continuous Sign Language Recognition by Iterative Training," in IEEE Transactions on Multimedia, vol. 21, no. 7, pp. 1880-1891, July 2019.

[2] S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02., Singapore, 2002, pp. 2204-2206 vol.5.

[3] Purohit, Kunal," A Wearable Hand Gloves Gesture Detection based on Flex Sensors for disabled People", 2017 International Journal for Research in Applied Science and Engineering Technology. V. 1825-1833. 10.22214/ijraset.2017.8261.

[4] A. S. Nikam and A. G. Ambekar, "Sign language recognition using image based hand gesture recognition techniques," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), Coimbatore, 2016, pp. 1-5.

[5] A. G. Jaramillo and M. E. Benalcázar, "Real-time hand gesture recognition with EMG using machine learning," 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Salinas, 2017, pp. 1-5.

[6] P. S. Rajam and G. Balakrishnan, "Real time Indian Sign Language Recognition System to aid deaf-dumb people," 2011 IEEE 13th International Conference on Communication Technology, Jinan, 2011, pp. 737-742.

[7] A. Saxena, D. K. Jain and A. Singhal, "Sign Language Recognition Using Principal Component Analysis," 2014 Fourth International Conference on Communication Systems and Network Technologies, Bhopal, 2014, pp. 810-813.

[8] H. Brashear, T. Starner, P. Lukowicz and H. Junker, "Using multiple sensors for mobile sign language recognition," Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings., White Plains, NY, USA, 2003, pp. 45-52.

[9] P. Mekala, Y. Gao, J. Fan and A. Davari, "Real-time sign language recognition based on neural network architecture," 2011 IEEE 43rd Southeastern Symposium on System Theory, Auburn, AL, 2011, pp. 195-199.

[10] Singha, Joyeeta & Das, Karen, "Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique", 2013 International Journal of Advanced Computer Science and Applications. 4. 10.14569/IJACSA.2013.040228.

[11] C. Liao, S. Su and M. Chen, "Vision-Based Hand Gesture Recognition System for a Dynamic and Complicated Environment," 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, 2015, pp. 2891-2895.