IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEVELOPING A GESTURE BASED VIDEO CALLING FOR DEAF AND MUTE PEOPLE USING MICROSOFT KINCET

R.Savithiri[1]

B.Priya[2] C.Kalaiarasi[3] K.Varalakshmi[4] V.Dharma Prakash[5]

Assistant Professor[1,2,3,4,5]

Department of Computer Science and Engineering

PERI INSTITUTE OF TECHNOLOGY

Abstract

In recent year, there has been rapid increase in the number of deaf-mute victims due to birth defects, accidents and oral diseases. Since deaf-mute people cannot communicate easily with normal people so they have to depend on some sort of visual communication. This system is based on a skin-color modelling technique so the skin color range is predetermined that will extract pixels(hand) from non-pixels(background). The images were fed into the model called the Convolutional Neural Network (CNN) for classification of images. Keras was used for training of images provided with proper lighting condition and a uniform background; the system acquired an average testing accuracy of 93.67%, of which 90.04% was attributed to ASL alphabet recognition and 97.52% for static word recognition, thus surpassing that of other related studies. The framework is developed by using python flask for a Web Based Application to establish the connection between deaf-mute and normal users. So that the Sign Language can be converted into Normal text or voice as output for the normal people. on the other end voice of the normal people is converted into text for the convenient of the deaf-mute people

Introduction

To Develop an Web Based Application for Two-Way Communication between Deaf-Mute users and Normal users by converting Sign language into text or voice message at the output with the help of deep learning algorithm we can achieve this operation for image processing and to build a better communication between any users. 1.2 Project Domain 1.2.1 Image processing Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing

technologies. It forms core research area within engineering and computer science disciplines too. Image processing basically includes the following three steps: Importing the image via image acquisition tools. Analysing and manipulating the image.Output in which result can be altered image or report that is based on image analysis.

Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital Technique: Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

Related works

A various hand gestures were recognized with different methods by different researchers in which were implemented in different fields. The recognition of various hand gestures were done by vision based approaches, data glove based approaches, soft computing approaches like Artificial Neural Network, Fuzzy logic, Genetic Algorithm and others like PCA, Canonical Analysis, etc. The recognition techniques are divided into three broad categories such as Hand segmentation approaches, Feature extraction approaches and Gesture recognition approaches. “Application research on face detection technology uses Open CV technology in mobile augmented reality” introduces the typical technology.

Open source computer vision library, Open CV for short is a cross-platform library computer vision based on open source distribution. The Open CV, with C language provides a very rich visual processing algorithm to write it part and combined with the characteristics of its open source. Data gloves and Vision based method are commonly used to interpret gestures for human computer interaction. The sensors attached to a glove that finger flexion into electrical signals for determining the hand posture in the data gloves method. The camera is used to capture the image gestures in the vision based method. The vision based method reduces the difficulties as in the glove based method. “Hand talk-a sign language recognition based on accelerometer and semi data” this paper introduces American Sign Language conventions. It is part of the “deaf culture” and includes its own system of puns, inside jokes, etc. It is very difficult to understand understanding someone speaking Japanese by English speaker. The sign language of Sweden is very difficult to understand by the speaker of ASL. ASL consists of approximately 6000 gestures of common words with spelling using finger used to communicate obscure words or proper nouns. “Hand gesture recognition and voiceconversion system for dumb people” proposed lower the communication gap between the mute community and additionally the standard world. The projected methodology interprets language into speech. The system overcomes the necessary time difficulties of dumb people and improves their manner. Compared with existing system the projected arrangement is simple as well as compact and is possible to carry to any places. This system converts the language in associate text into voice that's well explicable by blind and ancient people. The language interprets into some text kind displayed on the digital display screen, to facilitate the deaf people likewise. In world applications, this system is helpful for deaf and dumb of us those cannot communicate with ancient person

Methodology

In order to solve the feature extraction and classification of sign language, this paper proposes a new 3D CNN structure. Conv Blocks in sequence, and the input of each Conv Block is sampled to the same size as the output through the operation of 3D Average Pooling. Then they are concatenated with the feature map extracted by Conv Block to form new features, and input into the next Conv Block. After the output of Convblock5, the feature graph of 1×1×1 is obtained by using a Global Max Pooling, which passes through Dense Block, and get the probability of dichotomy by using Sigmoid function.

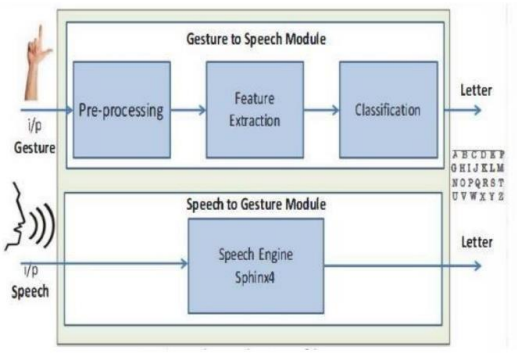


Figure 5.1 Architecture Diagram

Conv Block Structure

The sequence of Conv Block consists of 1 BN layer, activation function, 3D convolutional layer and 3D Max Pooling layer. The first four Conv Blocks use CReLU as the activation function, and the 5 Conv Block CReLUs replace Leaky ReLU, and do the operation of Pooling. Batch Normalization is usually located in front of the activation function. Through the normalization operation of each dimension of the data in Batch, it prevents saturated nonlinear function from gradient dispersion due to too large or too small input. At the same time, the normalization operation can prevent a large learning rate from causing a gradient explosion of back propagation, reduce the requirement of parameter initialization, and accelerate CNN training. The activation function is the CReLU function, that is: CReLU(x) = [ReLU(x), ReLU(−x)] (1) It doubles the input dimension, removes the convolution kernel of Pair-Grouping Phenomenon, reduces redundancy, and improves parameter utilization ratio. In the fifth Conv Block, it doesn't do the operation of 3D Max Pooling. The activation function is the Leaky ReLU, and takes the parameter a = 0.01, that is: Leaky ReLU(x) = max(x , ax) (2)

Dense Block Structure

Conv Block extracts features from images, while Dense Block acts as a classifier. The structure of Dense Block is shown in Fig.6, which consists of 2 Fully Connected Layer, 1 Batch norm layer and 1 Leaky ReLU activation function. Finally, it generates the confidence of binary classification by Sigmoid function.

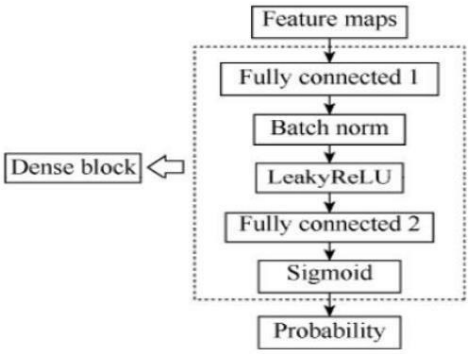


Fig.

Glove based approaches (Text Conversion) in this category requires signers to wear a sensor glove or a colored glove. The task will be simplified during segmentation process by wearing glove. The drawback of this approach is that the signer has to wear the sensor hardware along with the glove during the operation of the system.

Vision based approaches (Text Conversion) Image processing algorithms are used in Vision based technique to detect and track hand signs and facial expressions of the signer. This technique is easier to the signer since there is no need to wear any extra hardware. However, there are accuracy problems related to image processing algorithms and these problems are yet to be modified. There are again two different approaches in vision based sign language recognition: -3D model based -Appearance based 3D model based methods make use of 3D information of key elements of the body parts. Using this information, several important parameters, like palm position, joint angles etc., can be obtained. This approach uses volumetric or skeletal models, or a combination of the two. Volumetric method is better suited for computer animation industry and computer vision. This approach is very computational intensive and also, systems for live analysis are still to be developed. Appearance-based systems use images as inputs. They directly interpret from these videos/images. They don't use a spatial representation of the body. The parameters are derived directly from the images or videos using a template database. Some templates are the deformable 2D templates of the human parts of the body, particularly hands. The sets of points on the outline of an object called as deformable templates. It is used as interpolation nodes for the objects outline approximation

Figure 1. Gestures of sign recognition

Data acquisition

To achieve a high accuracy for sign recognition in sign language recognition system, 10 images will be taken for each 26 signs. These images are included in training and testing database. The captured image at a distance is adjusted by the signer to get the required image clarity.

Pre-processing

Pre-processing consist image acquisition, segmentation and morphological filtering methods.

Image acquisition

This is the first step of pre-processing. This is the process of sensing of an image. So in an image acquisition, image is sensed by "illumination". It will also involve pre-processing such as scaling. In image acquisition the image will be taken from database.

Segmentation

Segmentation is the process in which image is converted into small segments so that the more accurate image attribute can be extracted. If the segments are properly autonomous (two segments of an image should not have any identical information) then representation and description of image will be accurate and while taking rugged segmentation, the result will not be accurate. Here the Segmentation of hands is carried out to separate object and the background. Otsu algorithm is used for segmentation purpose. The segmented hand image is represented certain features. The following figure 4 shows the segmented of hand image.

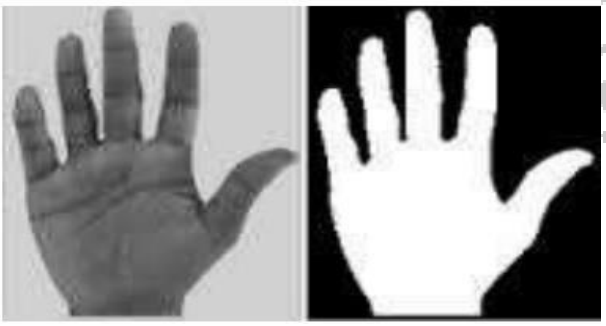


Fig.: Segmented of hand image

Feature extraction

The reduction of data dimensionality by encoding related information in a compressed representation and removing less discriminative data is called as Feature extraction Technique. Feature extraction is vital to gesture recognition performance. Therefore, the selection of which features to deal with and the extraction method are probably the most significant design decisions in hand motion and gesture recognition development. Here principal component is used as main features. The following **figure 7** explains the feature extraction method

Figure 7. Feature extraction method

Sign recognition

• Recognition Phase the following **figure 8** shows the dimensionality reduction technique.

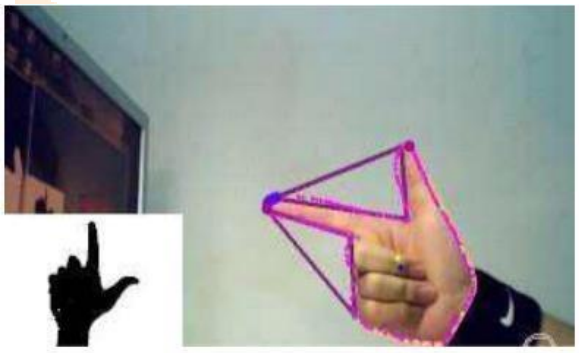
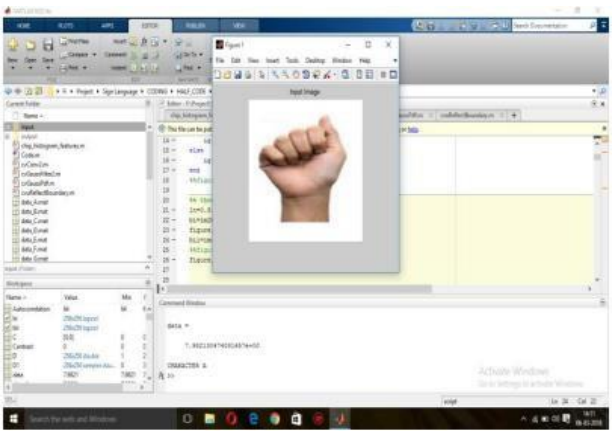


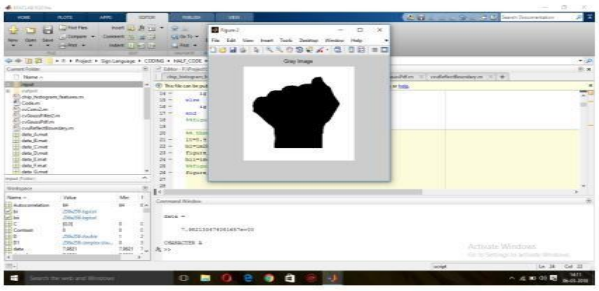
Figure 8. Dimensionality reduction

Each gesture is represented as a column vector in the training phase. These gesture vectors are then normalized with respect to average gesture. Next, the algorithm finds the eigenvectors of the covariance matrix of normalized gestures by using a speed up technique that reduces the number of multiplications to be performed. The corresponding gesture space projections were obtained by the eigenvector matrix then multiplied by each of the gesture vectors. In the recognition phase, a subject gesture is normalized with respect to the average gesture and then projected onto gesture space using the eigenvector matrix. Lastly, Euclidean distance is computed between this projection and all known projections. The minimum value of these comparisons is selected for recognition during the training phase. Finally, recognized sign is converted into appropriate text and voice which is displayed

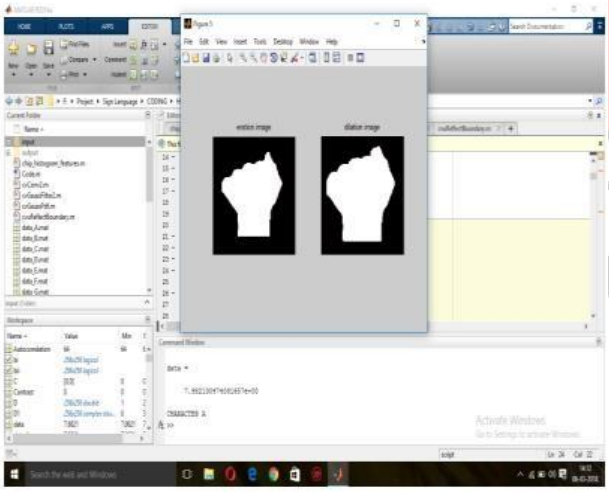
Experimental Results



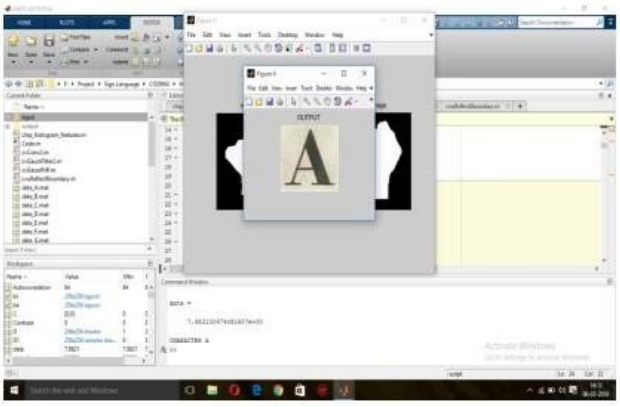
Input Image



Inversion of Grey Image



Erosion and Dilation



Output Image

Conclusions

The main objective of the project is to develop a system that can translate static sign language into its corresponding word equivalent that includes letters, numbers, and basic static signs to familiarize the users with the fundamentals of sign language. Reaching the training phase of the development, one of the objectives of the study was to match or even exceed the accuracy of the studies presented using deep learning. Our system was able to achieve 99% training accuracy, with testing accuracy of 90.04% in letter recognition, 93.44% in number recognition and 97.52% in static word recognition, obtaining an average of 93.667% based on the gesture recognition with limited time. Each system was trained using 2,400, 50 × 50 images of each letter/number/word

References

[1] F. R. Session, A. Pacific, and S. Africa, Senat'2117, 2014, pp. 1–3

[2] L. G. Zhang, Y. Chen, G. Fang, X. Chen, and W. Gao, "A vision-based sign language recognition system using tied-mixture density HMM," in Proc. the 6th International Conference on Multimodal Interfaces, 2004, pp. 198-204.

[3] Q. Wang, X. Chen, L. G. Zhang, C. Wang, and W. Gao, "Viewpoint invariant sign language recognition," Computer Vision and Image Understanding, vol. 108, no. 1-2, pp. 87–97, 2007.

[4] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1371-1375, 1998.

[5] C. Vogler and D. Metaxas, "Handshapes and movements: Multiple-channel American sign language handshapes and movements: Multiple-channel ASL recognition," in Proc. International Gesture Workshop, 2003, pp. 247-258.

[6] T. Chouhan, A. Panse, A. K. Voona, and S. M. Sameer, "Smart glove with gesture recognition ability for the hearing and speech impaired," in Proc. 2014 IEEE Global Humanitarian Technology Conference-South Asia Satellite (GHTC-SAS), 2014, pp. 105-110.

[7] M. Süzgün, H. Özdemir, N. Camgöz, A. Kındıroğlu, D. Başaran, C. Togay, and L. Akarun, "HospiSign: An interactive sign language platform for hearing impaired," Journal of Naval Sciences and Engineering, vol. 11, no. 3, pp. 75-92, 2015.

[8] J. A. Deja, P. Arceo, D. G. David, P. Lawrence, and R. C. Roque, "MyoSL: A Framework for measuring usability of two-arm gestural electromyography for sign language," in Proc. International Conference on Universal Access in Human Computer Interaction, 2018, pp. 146-159. 50

[9] C. Ong, I. Lim, J. Lu, C. Ng, and T. Ong, "Sign-language recognition through Gesture & Movement Analysis (SIGMA)," Mechatronics and Machine Vision in Practice, vol. 3, pp. 232-245, 2018.

N. Sandjaja and N. Marcos, "Sign language number recognition," in Proc. 2009 Fifth International Joint Conference on INC, IMS and IDC, 2009, pp. 1503- 1508.

[10] E. P. Cabalfin, L. B. Martinez, R. C. L. Guevara, and P. C. Naval, "Filipino sign language recognition using manifold learning," in Proc. TENCON 2012 IEEE Region 10 Conference, 2012,

pp. 1-5.

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

[12] J. P. Rivera and C. Ong, "Recognizing non-manual signals in Filipino sign language," in Proc. Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018, pp. 1-8.

[13] J. P. Rivera and C. Ong, "Facial expression recognition in Filipino sign language: Classification using 3D Animation units," in Proc. the 18th Philippine Computing Science Congress (PCSC 2018), 2018, pp. 1-8.

[14] J. Bukhari, M. Rehman, S. I. Malik, A. M. Kamboh, and A. Salman, "American sign language translation through sensory glove; SignSpeak," International Journal of u-and e-Service, Science and Technology., vol. 8, no. 1, pp. 131–142, 2015