Hand Gesture Recognition and Voice Conversion for Deaf and Dumb

¹R. Anusha, ²K. Dhanalakshmi, ³S. Shravanthi, ⁴G. Hymanjali, ⁵T. Hemalatha

1,2,3,4,5Department of Computer Science Engineering, Santhiram Engineering College, NH-40, Nerawada 'X' Roads, Nandyal, Kurnool-Dist, Andhra Pradesh, India - 518 501

ABSTRACT

Sign Language Recognition (SLR) targets on interpreting the sign language into text or speech, so as to facilitate the communication between deaf-mute people and ordinary people. This task has broad social impact, but is still very challenging due to complexity and large variations in hand actions. Existing methods for SLR use hand-crafted features to describe sign language motion and build classification models based on those features. However, it is difficult to design reliable features to adapt to large variations of hand gestures. To approach this problem, we propose a novel convolutional neural network (CNN) which extracts discriminative spatial-temporal features from raw video stream automatically without any prior knowledge, avoiding designing features. To boost the performance, multi-channels of video streams, including colour information, depth clue, and body joint positions, are used as input to the CNN in order to integrate colour, depth and trajectory information. We validate the proposed model on a real dataset collected with Microsoft Kinect and demonstrate its effectiveness over the traditional approaches based on hand-crafted features.

Keywords: Sign language, Hand gesture, Feature extraction, Gesture recognition, Principal Component Analysis

Introduction

DEEP LEARNING: Deep Learning is a subset of Machine Learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behaviour of human brain, though far from matching the ability, allows to learn from large amounts of data. Deep Learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have improved a lot in speech recognition, object recognition and detection and many other domains. Deep Learning discovers intricate shape in huge datasets by using the back propagation algorithm to indicate the change in internal parameters that are used for representation difference from each layer. Deep Convolutional networks are very helpful in processing the videos, images, speech, audio, which is different from another networks.

CONVOLUTIONAL NEURAL NETWORKS: Convolutional Neural Network is a Deep Learning algorithm, which is designed to process data that come in the form of multiple arrays. CNN can take input image, assign importance to **various** aspects in the image and be able to differentiate one from the other. CNN is most commonly used for visual imagery. Convolution Layers are great feature extractors for images. The Kernels in the convolution layers are learnable parameters just like neural networks parameters. Convolution Layers can do manual feature extraction. There are two types of layers in CNN, Convolutional Layer (Feature Extractor) and Fully Connected Layer (Classifier). Usually we use Pooling Layers in conjunction with the Convolution layers. In Convolutional Layers, We learn the Kernel weights. In Fully connected Layers, we learn the weights and biases. But in pooling layers, there are no parameters to learn, it is there to reduce the size of the feature maps.

OPENCV: OpenCV is a computer vision library, mainly used for developing real-time projects. OpenCV helps in understanding, reconstructing the 3Dimesional scene using 2Dimensional images. By using computer software and hardware, it tries to clone human vision. It can take input in either of the two ways, uploading image by the user or by using the camera(Video Capture). OpenCV helps in various applications like face recognization, face mask recognization and many more. OpenCV is a open source library, which can be used by any person for his/her applications. This library contains many useful functions and algorithms, which can be used freely

NUMPY: Numpy is a module for python. The name is an acronym for "Numeric python".Numpy enriches the programming language python with powerful data structures, implementing multi-dimensional arrays and matrices. These data structures generate efficient calculations with matrices and arrays**.**

Literature Review

In Literature Review, we studied about present task associated to this theme and attempt to recognize about the current gadget behavior. Shweta S. Shinde, Rajesh M. Autee and Vitthal K. Bhosale [1] have proposed a technique in which the attitude and height calculation strategy is used to extract the aspects of hand gestures by means of the usage of MATLAB and then they convert the identified gesture into speech the usage of MATLAB built in command.

Sangeetha. R. K, Valliammai. V and Padmavathi. S [2] have proposed a gadget primarily based on the Indian hand signal language which carries each fingers to create a gesture in contrast to the American signal language in which one hand is used. Their gadget is applied the use of MATLAB barring the use of any different exterior hardware for the user, right here the runtime stay photograph is captured after which photograph frames are extracted and picture processing is utilized the use of HIS mannequin and then the characteristic extraction is finished by means of distance seriously change method. The effects got by means of this mannequin is determined to be satisfactorily right for most of the hand signs.

Existing System

Sign Language is the imperative strategies for correspondence for difficult of listening to people. With motion of science and improvement more than a few frameworks have been made now not solely to avert the problem of difficult of listening to and bonehead humans but however execute it in extraordinary fields.Sign language is a language which as adversarial to voice or sound fashions makes use of guide correspondence and non -verbal correspondence to pass by on the message. This accommodates normally the combination of shapes, presentation and development of the hands

Proposed System

In our proposed system, we propose a novel convolutional neural network(CNN) which extracts discriminative spatial-temporal features from raw video stream automatically without any prior knowledge, avoiding designing features. Technically spealing, the main challenge of sign language recognition lies in developing descriptors to express hand-shapes and motion trajectory. In particular, handshape description involves tracking hand regions in video stream, segmenting hand-shape images from complex background in each frame and gestures recognition problems. Motion trajectory is also related to tracking of key points and curve matching. Although lots of research works have been conducted on these two issues for now, it is sill hard to obtain satisfying result for SLR due to the variation and occlusion of hands and body joints.

Kinect is a action sensor which can supply colorstream. With the public home windows SDK, the physique joint locations can be acquired in real-time. Therefore, wechoose Kinect as seize system to report signal words dataset. The alternate of coloration and depth in pixel level are beneficial records to discriminate unique sign actions. And the variant of physique joints in time dimension can depict the trajectory of signal actions. Using more than one kinds of visual sources as enter leads CNNs paying interest to alternate no longer solely in color, but also in depth and trajectory. It is well worth bringing up that we can keep away from the issue of monitoring hands, segmenting fingers from historical past and designing descriptors for arms due to the fact CNNs have the capability to study facets mechanically from raw data besides any prior knowledge.

CNNs have been utilized in video circulation classification recently years. A manageable challenge of CNNs is time consuming. It prices countless weeks or months to instruct a CNNs with million-scale in million videos. Fortunately, it is nevertheless viable to reap real-time efficiency, with the assist of CUDA for parallel processing.
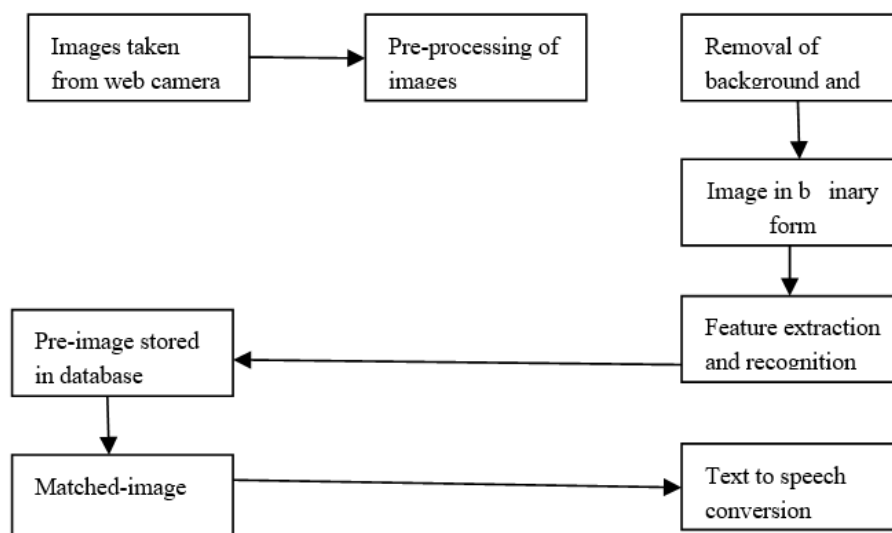
Working Principle

Image Matching Process

Image matching is a system in which the captured photo is in contrast with the pictures saved in database

Feature Extraction

Principal factor evaluation (PCA) is one of the statistical methods often used in sign processing to the statistics dimension discount or to the information decorrelation. Presented paper offers with two wonderful purposes of PCA in photo processing. The first software consists in the photo coloration discount whilst the three colour aspects are decreased into one containing a foremost section of information. The 2d use of PCA takes benefit of eigenvectors residences for willpower of chosen object orientation. Various strategies can be used for preceding object detection. Quality of picture segmentation implies to consequences of the following procedure of object orientation comparison primarily based on PCA as well. Presented paper quickly introduces the PCA idea and Results are documented for the chosen actual pictures. Principal thing evaluation (PCA) belongs to linear transforms based totally on the statistical techniques

Classification Method

K-Nearest Neighbor is a Learning algorithm that Defer in the selection to generalize past the coaching examples until a new question is encountered. Whenever we have a new factor to classify, classify, we discover its K nearest neighbors from the coaching data. The distance is calculated the use of n Euclidean Distance. "Support Vector Machine" (SVM) is a supervised desktop studying algorithm which can be used for each classification and regression challenges. However, it is primarily used in classification problems. In this algorithm, we plot every records object as a factor in n-dimensional house (where n is wide variety of points you have) with the fee of every characteristic being the fee of a specific coordinate. Then, we operate classification by way of discovering the hyper-plane that differentiate the two training very well. Support Vectors are absolutely the co-ordinates of character observation. Support Vector Machine is a frontier which exceptional segregates the two classes.

Results

Database creation   Image Capturing   Final Result

Database can be growing correctly by way of person which is used for coaching the system, the solely trouble whilst growing database is that history have to white and clear. The person has to specify the variety of samples; wide variety of pattern essentially capability quantity of photos in database.

Conclusion

We developed a CNN mannequin for signal language recognition. Our mannequin learns and extracts each spatial and temporal points by means of performing 3D convolutions. The developed deep structure extracts multiple types of records from adjoining enter frames and then performs convolution and sub-sampling separately. The remaining characteristic illustration combines information from all channels. We use multilayer perceptron classifier to classify these feature representations. For comparison, we consider both CNN and GMM-HMM on the identical dataset. The experimental effects exhibit the effectiveness of the proposed method.

References

[1]. Shinde, Shweta S., Rajesh M. Autee, and Vitthal K. Bhosale. "Real time two way communication approach for hearing impaired and dumb person based on image processing." Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on. IEEE, 2016.

[2]. Shangeetha, R. K., V. Valliammai, and S. Padmavathi. "Computer vision based approach for Indian Sign Language character recognition." Machine Vision and Image Processing (MVIP), 2012 International Conference on. IEEE, 2012.

[3]. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105

[4]. Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei- Fei, "Large-scale video classification with convolutional neural networks," in CVPR, 2014. [3] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick ´Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[5]. Hueihan Jhuang, Thomas Serre, Lior Wolf, and Tomaso Poggio, "A biologically inspired system for action recognition," in Computer Vision,

2007. ICCV 2007. IEEE 11th International Conference on. Ieee, 2007, pp. 1–8. [5] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, "3D convolutional neural networks for human action recognition," IEEE TPAMI, vol. 35, no. 1, pp. 221–231, 2013

[6].   Kirsti Grobel and Marcell Assan, "Isolated sign language recognition using hidden markov models," in Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on. IEEE, 1997, vol. 1, pp. 162–167