Recognizing Sign Language using Convolutional Neural Networks

S.Varalakshmi[1]|CH.Sangeetha[2]|K.Lakshmi Kalpana[3]|Nenduguri Sai Nithin[4]

1 ,2 & 3 Assistant Professor, CSE department, Kasireddy Narayanreddy College of Engineering And Research, Hyderabad, TS.

4 UG SCHOLAR, CSE department, Kasireddy Narayanreddy College of Engineering And Research, Hyderabad, TS.

ABSTRACT: The goal of Sign Language Recognition (SLR) is to translate sign language into speech or text in order to let deaf-mute people and regular people communicate. Although this exercise has a wide-ranging social influence, its complexity and wide range of hand motions make it extremely difficult. Current SLR techniques characterize sign language motion using manually created characteristics, then use those features to create classification models. Designing dependable features that can adjust to the wide range of hand movements is challenging, though. In order to tackle this issue, we suggest a brand-new convolution neural network (CNN) that automatically and without prior knowledge extracts discriminative spatial-temporal characteristics from raw video streams, avoiding feature creation. Multi-channel video streams with color information, depth clues, and body joint locations are fed into the CNN to integrate color, depth, and trajectory data in order to improve performance. We show the efficacy of the suggested model above the conventional methods based on manually created features by validating it on an actual dataset gathered using a Microsoft Kinect.

KEYWORDS: SLR, CNN, Deaf-mute people, adapt.

I.INTRODUCTION:

Sign language, as one of the most widely used communication means for hearing-impaired people, is expressed by variations of hand-shapes, body movement, and even facial expression. Since it is difficult to collaboratively exploit the information from hand-shapes and body movement trajectory, sign language recognition is still a very challenging task. This paper proposes an effective recognition model to translate sign language into text or speech in order to help the hearing impaired communicate with normal people through sign language.

Technically speaking, the main challenge of sign language recognition lies in developing descriptors to express hand-shapes and motion trajectory. In particular, hand-shape description involves tracking hand regions in video stream, segmenting hand-shape

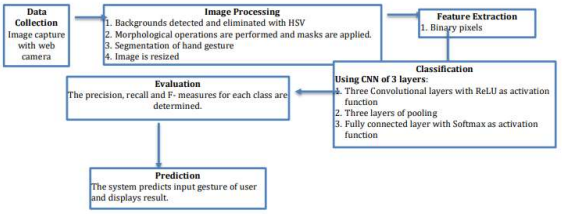
images from complex background in each frame and gestures recognition problems. Motion trajectory is also related to tracking of the key points and curve matching. Although lots of research works have been conducted on these two issues for now, it is still hard to obtain satisfying result for SLR due to the variation and occlusion of hands and body joints. Besides, it is a nontrivial issue to integrate the hand-shape features and trajectory features together. To address these difficulties, we develop a CNNs to naturally integrate hand-shapes, trajectory of action and facial expression. Instead of using commonly used color images as input to networks like [1, 2], we take color images, depth images and body skeleton images simultaneously as input which are all provided by Microsoft Kinect.

Kinect is a motion sensor which can provide color stream and depth stream. With the public Windows SDK, the body joint locations can be obtained in real-time as shown in Fig.1. Therefore, we choose Kinect as capture device to record sign words dataset. The change of color and depth in pixel level are useful information to discriminate different sign actions. And the variation of body joints in time dimension can depict the trajectory of sign actions. Using multiple types of visual sources as input leads CNNs paying attention to the change not only in color, but also in depth and trajectory. It is worth mentioning that we can avoid the difficulty of tracking hands, segmenting hands from background and designing descriptors for hands because CNNs have the capability to learn features automatically from raw data without any prior knowledge [3].

CNNs have been applied in video stream classification recently years. A potential concern of CNNs is time consuming. It costs several weeks or months to train a CNNs with million-scale in million videos. Fortunately, it is still possible to achieve real-time efficiency, with the help of CUDA for parallel processing. We propose to apply CNNs to extract spatial and temporal features from video stream for Sign Language Recognition (SLR). Existing methods for SLR use hand-crafted features to describe sign language motion and build classification model based on these features. In contrast, CNNs can capture motion information from raw video data automatically, avoiding designing features. We develop CNNs taking multiple types of data as input. This architecture integrates color, depth and trajectory information by performing convolution and sub sampling on adjacent video frames. Experimental

results demonstrate that 3D CNNs can significantly outperform Gaussian mixture model with Hidden Markov model (GMM-HMM) baselines on some sign words recorded by ourselves.

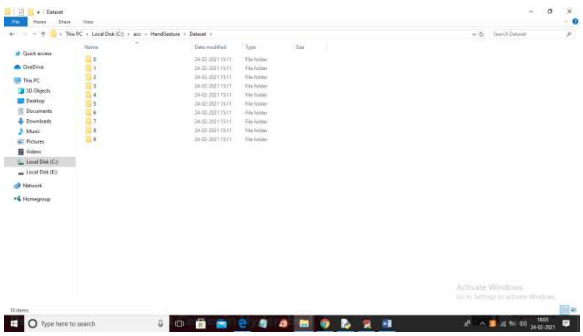
II.PROPOSED SYSTEM: The first step of the proposed system is to collect data. Many researchers have used sensors or cameras to capture the hand movements. For our system, we make use of the web camera to shoot the hand gestures. The images undergo a series of processing operations whereby the backgrounds are detected and eliminated using the colour extraction algorithm HSV(Hue,Saturation,Value). Segmentation is then performed to detect the region of the skin tone. Using the morphological operations, a mask is applied on the images and a series of dilation and erosion using elliptical kernel are executed . With open CV, the images obtained are amended to the same size so there is no difference between images of different gestures . Our dataset has 2000 American sign gesture images out of which 1600 images are for training and the rest 400 are for testing purposes. It is in the ratio 80:20. Binary pixels are extracted from each frame, and Convolutional Neural Network is applied for training and classification. The model is then evaluated and the system would then be able to predict the alphabets.



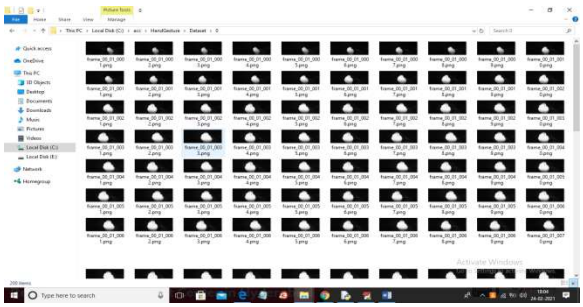
Data collection is indelibly an essential part in this research as our result highly depends on it. We have therefore created our own dataset of ASL having 2000 images of 10 static alphabet signs. We have 10 classes of static alphabets which are A,B,C,D,K,N,O,T and Y. Two datasets have been made by 2 different signers. Each of them has performed one alphabetical gesture 200 times in alternate lighting conditions. The dataset folder of alphabetic sign gestures is further split into 2 more folders, one for training and the other for testing. Out of the 2000 images captured, 1600 images are used for training and the rest for testing. To get higher consistency, we have captured the photos in the same background with a webcam each time a command is given. The images obtained are saved in the png format .It is to be pinpointed that there is no loss in quality whenever an image in png format is opened ,closed and stored again.PNG is also good in handling high contrast and detailed image. The webcam will capture the images in the RGB color space.

III.RESULTS:

In this project using CNN we are recognizing hand gesture movement and to train CNN we are using following images shown in below screen shots



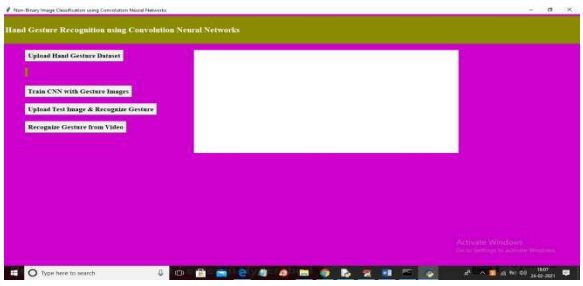
In above screen we can see we have 10 different types of hand gesture images and to see those images just go inside any folder



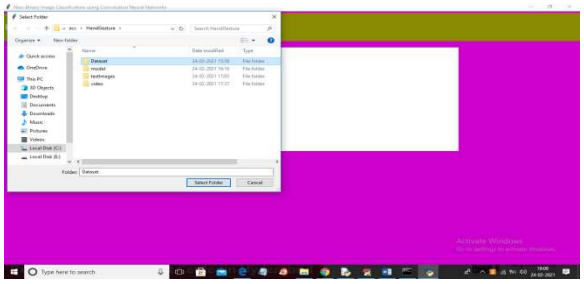
In above screen showing images from 0 folder and similarly you can see different images in different folders.

SCREEN SHOTS

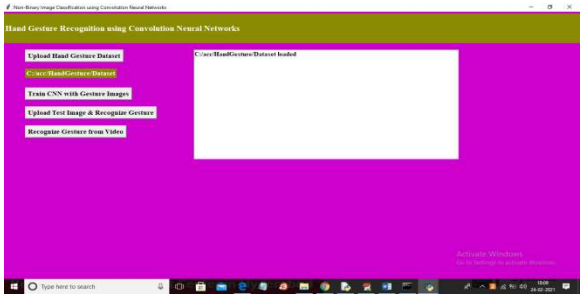
To run project double click on run.bat file to get below screen



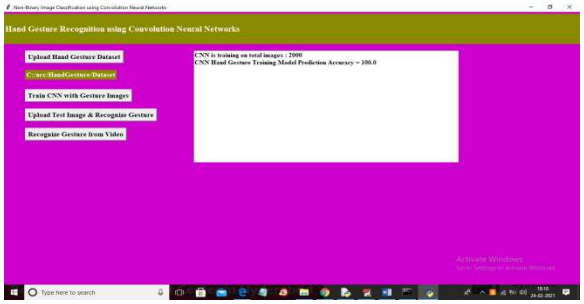
In above screen click on 'Upload Hand Gesture Dataset' button to upload dataset and to get below screen



In above screen selecting and uploading 'Dataset' folder and then click on 'Select Folder' button to load dataset and to get below screen

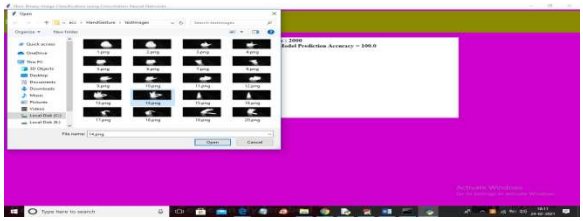


In above screen dataset loaded and now click on 'Train CNN with Gesture Images' button to trained CNN model and to get below screen

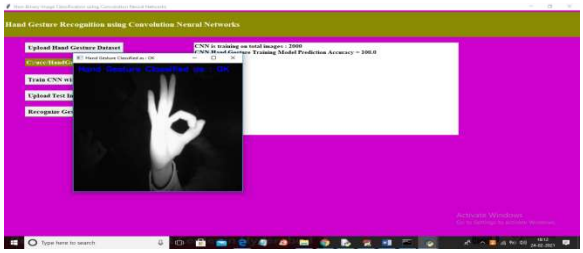


In above screen CNN model trained on 2000 images and its prediction accuracy we got as 100% and now model is ready and now click on 'Upload Test Image & Recognize

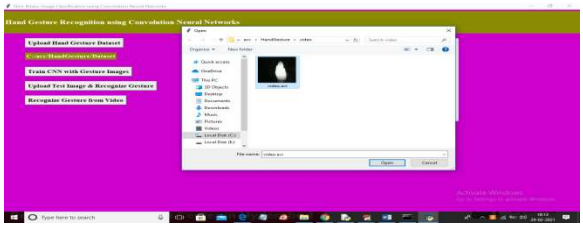
Gesture' button to upload image and to gesture recognition



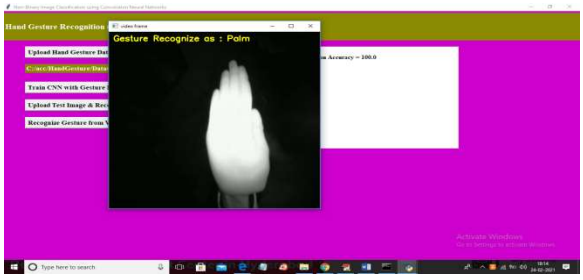
In above screen selecting and uploading '14.png' file and then click Open button to get below result

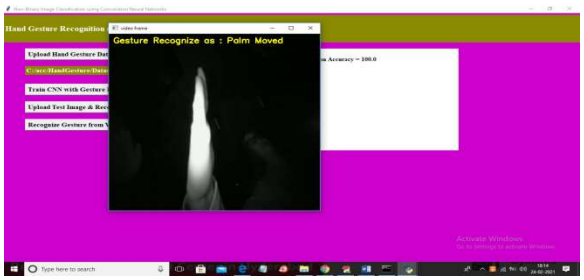


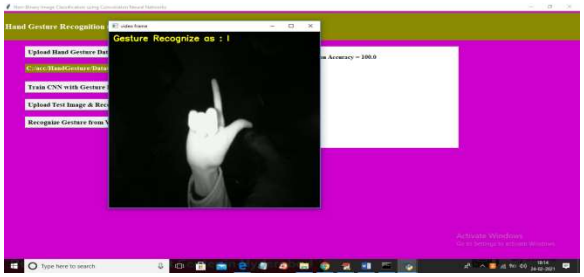
In above screen gesture recognize as OK and similarly you can upload any image and get result and now click on 'Recognize Gesture from Video' button to upload video and get result



In above screen selecting and uploading 'video.avi' file and then click on 'Open' button to get below result







In above screen as video play then will get recognition result

IV.CONCLUSION:

We created a CNN model to recognize sign language. Our model uses 3D convolutions to learn and extract temporal and spatial features. The designed deep architecture conducts convolution and sub-sampling independently after extracting various kinds of information from neighboring input frames. All of the channels' information is combined in the final feature representation. To classify these feature representations, we employ a multilayer perceptron classifier. For comparison, we use the same dataset to assess CNN and GMM-HMM. The outcomes of the experiment show how effective the suggested approach is.

REFERENCES

[1] 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.

[2] 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.

[4] 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.

[7] Thad Starner, Joshua Weaver, and Alex Pentland, “Realtime american sign language recognition using desk and wearable computer based video,” IEEE TPAMI, vol. 20, no. 12, pp. 1371–1375, 1998.

[8] Christian Vogler and Dimitris Metaxas, “Parallel hidden markov models for american sign language recognition,” in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. IEEE, 1999, vol. 1, pp. 116–122.

[9] Kouichi Murakami and Hitomi Taguchi, “Gesture recognition using recurrent neural networks,” in Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1991, pp. 237–242.

[10] Chung-Lin Huang and Wen-Yi Huang, “Sign language recognition using model-based tracking and a 3D hopfield neural network,” Machine vision and applications, vol. 10, no. 5-6, pp. 292–307, 1998.