Journal of Science and Technology

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

www.jst.org.in

DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57

SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

DR.N.SREEKANTH[1], PRACHI[2], K.SAI KIRTHANA[3], K. KEERTHANA[4], G. HARSHITHA[5]

[1]Professor, Department of ECE Malla Reddy Engineering College for Women (UGC-Autonomous) Maisammaguda,Hyderabad-500100

[2,3,4,5] UG Students, Department of ECE Malla Reddy Engineering College for Women (UGC-Autonomous) Maisammaguda,Hyderabad-500100

To Cite this Article

Swarnima Singh and Sunita Mishra, "Empower Teenagers and Farmers about Food Preservation and Processing Technique" *Journal of Science and Technology, Vol. 08, Issue 07,-July 2023, pp47-57*

Article Info

Received: 13-07-2023 *Revised: 15-07-2023* *Accepted: 20-07-2023* *Published: 24-07-2023*

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 the 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 the 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 color information, depth clue, and body joint positions, are used as input to the CNN in order to integrate color, 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

Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57

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,

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57

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 a CNNs taking multiple types of data as input. This architecture integrates color, depth and trajectory information by performing convolution and subsampling 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.

RELATED STUDY

As well stipulated by Nelson Mandela[1], Talk to a man in a language he understands, that goes to his head. Talk to him in his own language, that goes to his heart, language is undoubtedly essential to human interaction and has existed since human civilisation began. It is a medium humans use to communicate to express themselves and understand notions of the real world. Without it, no books, no cell phones and definitely not any word I am writing would have any meaning. It is so deeply embedded in our everyday routine that we often take it for granted and dont realise its importance. Sadly, in the fast changing society we live in, people with hearing impairment are usually forgotten and left out. They have to struggle to bring up their ideas, voice out their opinions and express themselves to people who are different to them. Sign language, although being a medium of commu- nication to deaf people,

Journal of Science and Technology

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

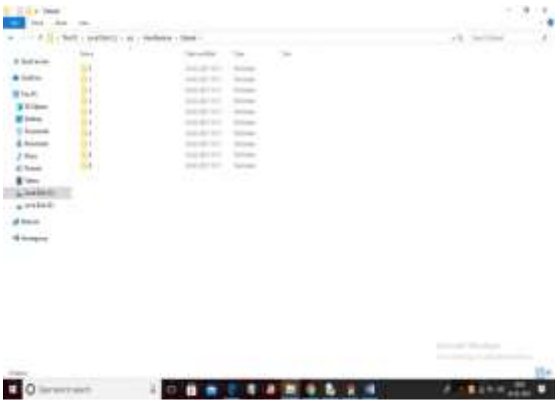
www.jst.org.in

DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57

still have no meaning when conveyed to a non-sign language user. Hence, broadening the communication gap.To prevent this from happening, we are putting forward a sign language recognition system. It will be an ultimate tool for people with hearing disability to communicate their thoughts as well as a very good interpretation for non sign language user to understand what the latter is saying. Many countries have their own standard and interpretation of sign gestures. For instance, an alphabet in Korean sign language will not mean the same thing as in Indian sign language. While this highlights diversity, it also pinpoints the complexity of sign languages. Deep learning must be well versed with the gestures so that we can get a decent accuracy. In our proposed system, American Sign Language is used to create our datasets. Figure 1 shows the American Sign Language (ASL) alphabets.

PROPOSED SYSTEM EXPLANATION

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



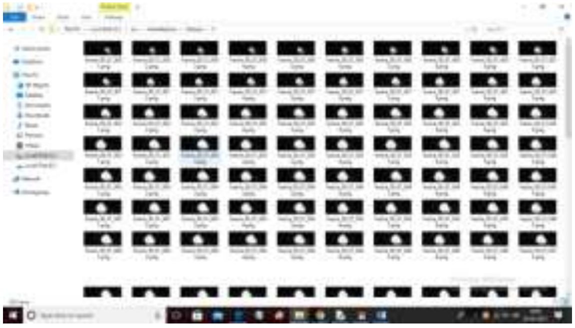
In above screen we can see we have 10 different types of hand gesture images and to see those images just go inside any folder

Journal of Science and Technology

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

www.jst.org.in

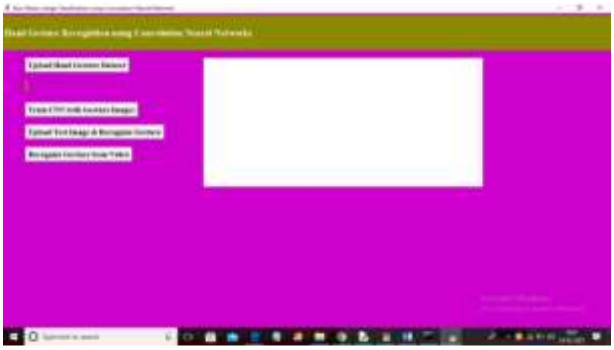
DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57



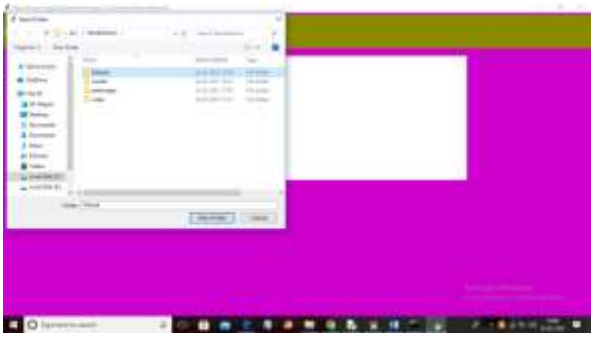
In above screen showing images from 0 folder and similarly you can see different images in different folders.

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



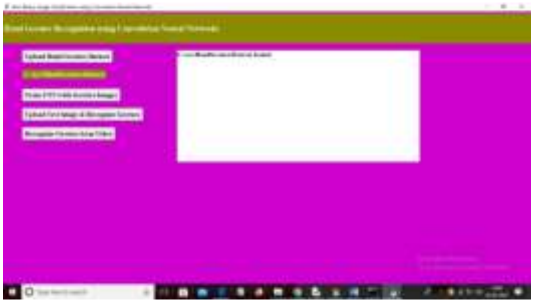
In above screen selecting and uploading 'Dataset' folder and then click on 'Select Folder' button to load dataset and to get below screen

Journal of Science and Technology

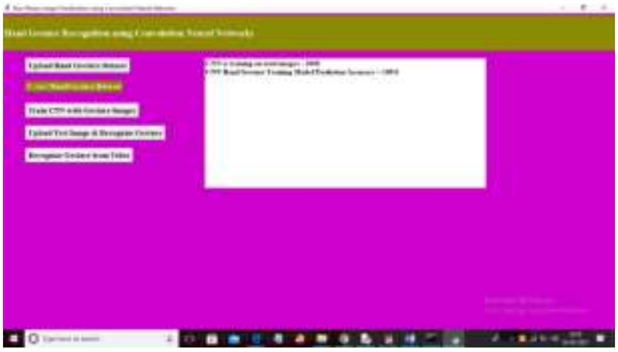
ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

www.jst.org.in

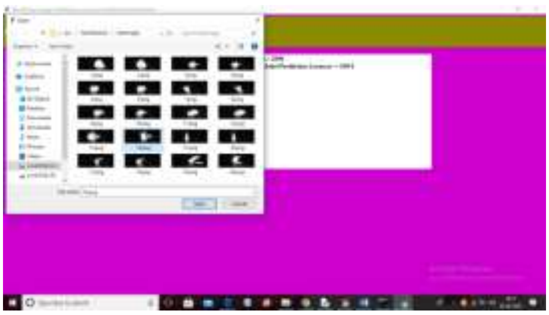
DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57



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



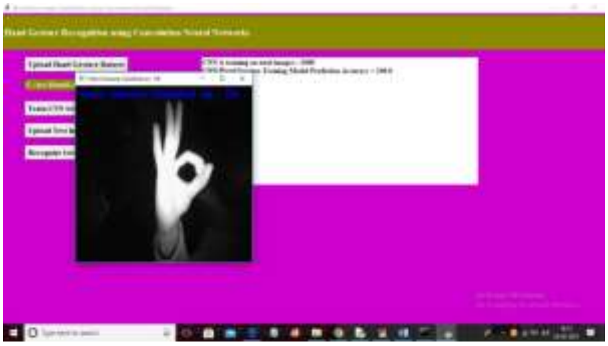
In above screen selecting and uploading '14.png' file and then click Open button to get below result

Journal of Science and Technology

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

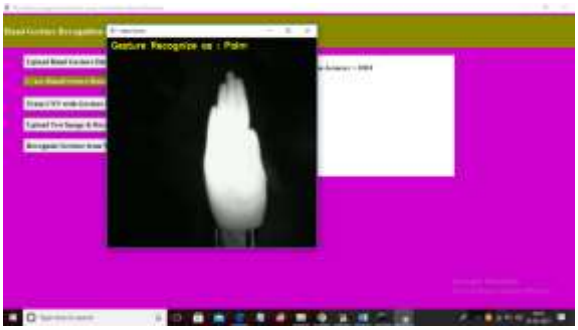
www.jst.org.in

DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57

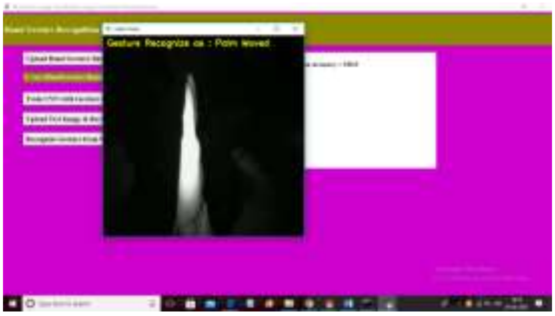


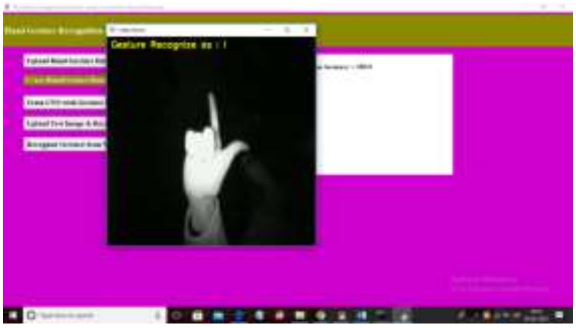
In above screen gesture recognize as OK and similarly you can upload any image and get result and now click on 'Recognize Gesture from Video' button to upload video and get result

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



Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)
www.jst.org.in
DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57





In above screen as video play then will get recognition result

CONCLUSION

We developed a CNN model for sign language recognition. Our model learns and extracts both spatial and temporal features by performing 3D convolutions. The developed deep architecture extracts multiple types of information from adjacent input frames and then performs convolution and subsampling separately. The final feature representation combines information from all channels. We use multilayer perceptron classifier to classify these feature representations. For comparison, we evaluate both CNN and GMM-HMM on the same dataset. The experimental results demonstrate the effectiveness of the proposed method.

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.

Journal of Science and Technology

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

www.jst.org.in

DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57

[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.

Journal of Science and Technology

ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)

www.jst.org.in

DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp47 - 57