Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

SIGN LANGUAGE RECOGNITION TO TEXT AND VOICE CONVERSION USING CNN

Ms. CHAGANTI VAISHNAVI[1], Ms. ELATI PRASANTHI[2],Ms.KOTAGIRI AISWARYA[3],Ms.DHUDALA SAI DURGA[4], Mrs. Dr ACP.RANJANI[5]

1. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India. Email id: chagantivaishnavi@gmail.com
2. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India.
3. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India.
4. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India.
5. Associate Professor, Computer Science and Engineering, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India. Email id: ranjani.vvnk@gmail.com

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

INDEX TERMS - convolutional neural network**,** Sign Language Recognition, hand gestures

1. 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].

During the last three decades, transform-based image compression technologies have been extensively researched and some standards have appeared. Two most common options of transformation are the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT). The DCT-based encoder can be thought of as compression of a stream of 8 × 8 small block of images. This transform has been adopted in JPEG. The JPEG compression scheme has many advantages such as simplicity, universality and availability. However, it has a bad performance at low bit-rates mainly because of the underlying block-based DCT scheme.

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 CNN can capture 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.

2. LITERATURE SURVEY

- An Approach for Minimizing the Time Taken for Translating Sign Language to Simple Sentence in English" Aradhana Kar, Pinaki Sankar Sign Language is the language of deaf. Thereare different types of sign languages spread all over the world. American Sign Language (ASL) is one of the sign languages. ASL is used by deaf Americans. We had created a system that translates sign language videos to simple sentences in English.

- "Deep Convolutional Neural Networks for Sign Language Recognition" G.Anantha Rao, Guntur (DT) Extraction of complex head and hand movements along with their constantly changing shapes for recognition of sign language is considered a difficult problem in computervision

- "American Sign Language Recognition using Deep Learning and Computer Vision" Kshitij Bantupalli, Ying Xie Speech impairment is a disability which affects an

individual's ability to communicate using speech and hearing. People who are affected by this use of other media or communication such as sign language.

- "Recent Developments in Sign Language Recognition Systems" M.F. Tolba, A.S.Elons Automated translation systems for sign languages are important in a world that is showing a continuously increasing interest in removing barriers faced by physically challenged individuals in communicating and contributing to the society and the workforce.

- "Interactive Systems for Sign Language Learning" Iurii Krak, ii Kryvonos In the article theproblems of communication of deaf people using sign language are considered. An analysis ofsign language information transfer which includes human hands, body, fingers movements, change of mimicry and emotions on human face is brought.

- "Moment Based Sign Language Recognition For Indian Languages" Umang Patel, Aarti G. Ambekar Communication plays a major role in day to day life. But it is very difficult for normal people to communicate with deaf, dumb & blind people & vice versa.

- "Hand Sign Language Recognition for Bangla Alphabet using Support Vector Machine" Md Azher Uddin, Shayhan Ameen Chowdhury The sign language considered as the main languagefor deaf and dumb people. So, a translator is needed when a normal person wants to talk with adeaf or dumb person. In this paper, we present a framework for recognizing Bangla Sign Language (BSL) using Support Vector Machine.

- "Sign Language Learning System with Image Sampling and Convolutional Neural Network" Yangho Ji, Sunmok Kim, Ki-Baek Lee This paper proposes a novel sign language learning system based on 2D image sampling and concatenating to solve the problems of conventional sign recognition. The system constructs the training data by sampling and concatenating from a sign language demonstration video at a certain sampling rate.

- "Machine Learning Techniques for Indian Sign Language Recognition" Kusumika Krori Dutta, Sunny Arokia Swamy Bellary Sign language is the only medium through which especially abled people can connect to the rest of the world through different hand gestures. With the advances in machine learning techniques, Hand gesture recognition (HGR) became a very important research topic.

- "Gesture Recognition Using Kinect for Sign Language Translation" Harsh Vardhan Verma, Eshan Aggarwal, Satish Chandra Sign Language is a widely used method of communication among the community of deaf-mute people. It contains some series of body gestures, which enables a person to interact without the need of spoken words. Although the use of sign languageis very popular among the deaf-mute people but the other communities don't even try to learn it, this creates a gulf of communication and hence becomes a cause of the isolation of physically impaired people

3. IMPLEMENTATION STYDY

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 express hand-shapes and motion trajectory. In particular, hand-

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

shape description involves tracking hand regions in video stream, segmenting hand-shape images from complex background in each frame and gestures recognition problems.

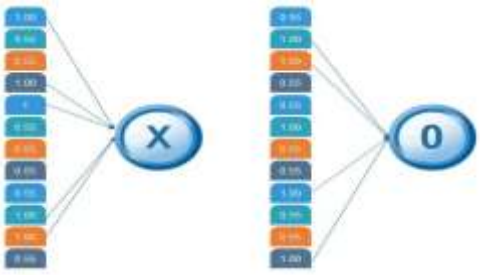
3.1 Proposed Methodology

In this article, the filtering of images plays an important role. It improves the accuracy of identifying the symbols even in low light areas. Before the process of saturation and grey scaling the image is sent to the filtering system where it tries to find the symbol shown in the hands, after recognizing the symbol the image is further processed and final result which is the word is obtained

3.2 Methodology

When we feed in, '**X**' and '**0**'. Then there will be some element in the vector that will be high. Consider the image below, as we can see for 'X' there are different top elements, and similarly, for 'O' we have various high elements.

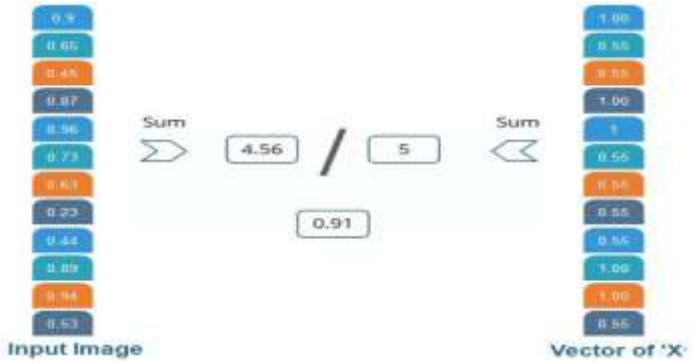
There are specific values in my list, which were high, and if we repeat the entire process which we have discussed for the different individual costs. Which will be higher, so for an **X** we have 1st, 4^{th}, 5^{th}, 10^{th}, and the 11^{th} element



of vector values are higher. And for **O** we have 2^{nd}, 3^{rd}, 9^{th} and 12^{th} element vector which are higher. We know now if we have an input image which has a 1^{st}, 4^{th}, 5^{th}, 10^{th}, and 11^{th} element vector values high. We can classify it as X similarly if our input image has a list which has the 2^{nd} 3^{rd} 9^{th} and 12^{th} element vector values are high so that we can organize it.

Comparing the Input Vector with X

After the training is done the entire process for both 'X' and 'O.' Then, we got this 12 element vector it has 0.9, 0.65 all these values then now how do we classify it whether it is **X** or **O**. We will compare it with the list of X and O so we have got the file in the previous slide if we notice we have got two different lists for X and O. We are comparing this new input image list that we have arrived with the X and O. First let us compare that with X now as well for X there are certain values which will be higher and nothing but 1st 4^{th} 5^{th} 10^{th} and 11^{th} value. So, we are going to sum them, and we have got 5= 1+ 1+ 1+ 1+1 times 1 we got 5, and we are going to sum the corresponding values of our image vector. So the 1^{st} value is 0.9 then the 4^{th} value is 0.87 5^{th} value is 0.96, and 10^{th} value is 0.89, and 11^{th} value is 0.94 so after doing the sum of these values have got
4.56 and divide this by 5 we got **0.9**.



We are comparing the input vector with 0.

And for X then we are doing the same process for O we have notice 2^{nd} 3^{rd} 9^{th}, and 12^{th} element vector values are high. So when we sum these values, we get 4 and when we do the sum of the corresponding values of our input image. We have got 2.07 and when we divide that by 4 we got **0.51**.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

4. RESULTS and EVOLUTION METRICS

Fig 1: 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

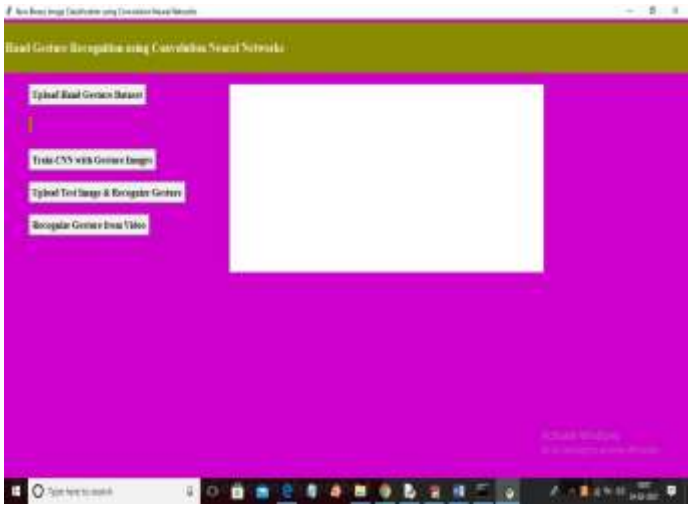


Fig2: In below screen click on 'Upload Hand Gesture Dataset' button to upload dataset and to getbelow

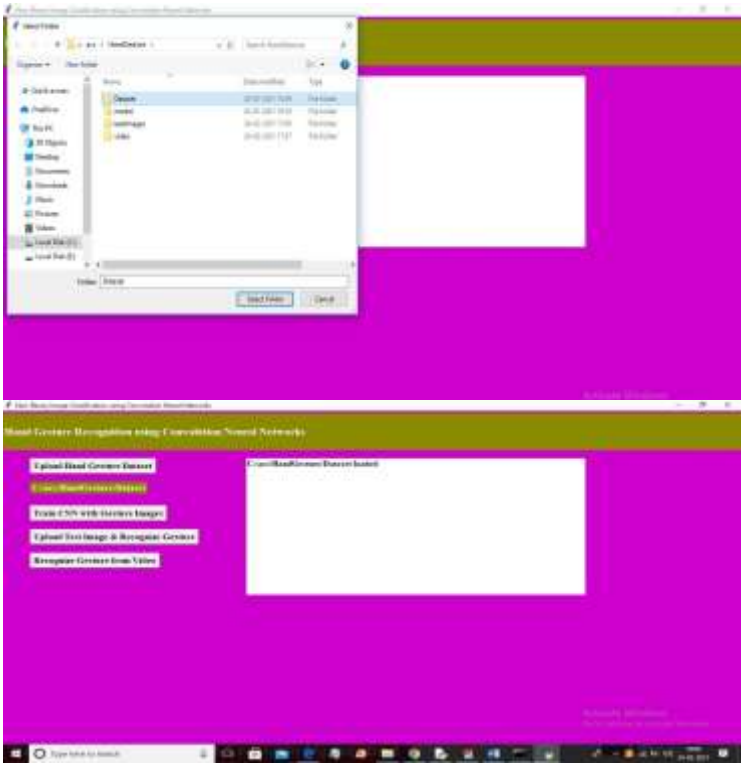
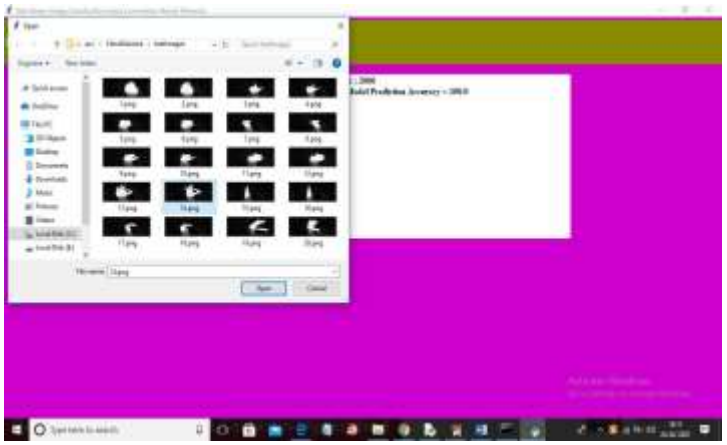


Fig 3: In above screen selecting and uploading 'Dataset' folder and then click on 'Select Folder' buttonto load dataset and to get below screen

Fig 4: In above screen dataset loaded and now click on 'Train CNN with Gesture Images' button to trained CNN model and to get below screen



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

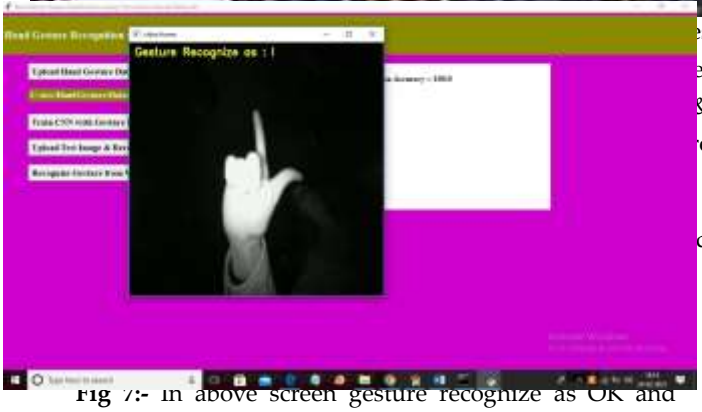


Fig 7:- In above screen gesture recognize as OK and similarly you can upload any image and get result and now click on 'Recognize Gesture from Video' button to upload video and get result

.

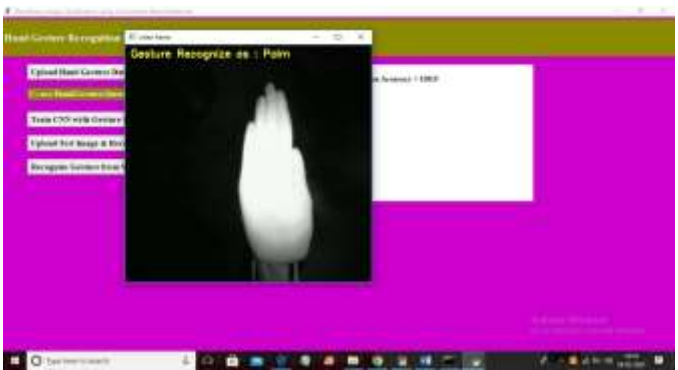
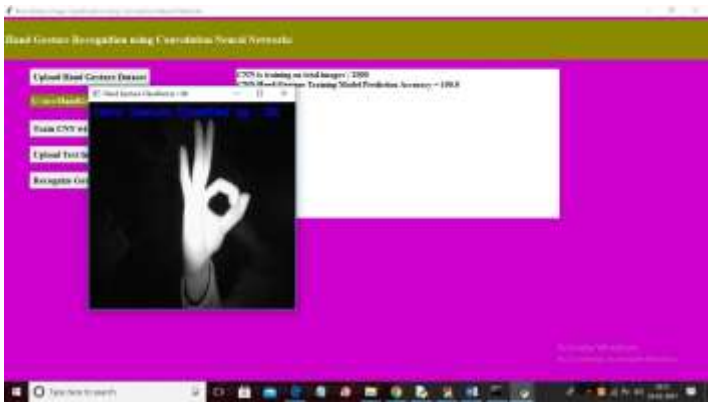


Fig 8 : In above screen as video play then will get recognition result

Fig 9 : In above screen as video play then will get recognition result

5. CONCLUSION

Nowadays, applications need several kinds of images as sources of information for elucidation and analysis. Several features are to be extracted so as to perform various applications. When an image is transformed from one form to another such as digitizing, scanning, and communicating, storing, etc. degradation occurs. Therefore ,the output image has to undertake a process called image enhancement, which contains of a group of methods that seek to develop the visual presence of an image. Image enhancement is fundamentally enlightening the interpretability or awareness of information in images for human listeners and providing better input for other automatic image processing systems. Image then undergoes feature extraction using various methods to make



the image more readable by the computer.Sign language recognition system is a powerful tool to preparae an expert knowledge, edge detect and the combination of inaccurate information from different sources the intend of convolution neural network is to get the appropriate classification.

6. REFERENCES

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.

1) 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.

2) 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.

3) 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.

4) 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.

5) 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.

6) 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.

7) 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.

8) Jong-Sung Kim, Won Jang, and Zeungnam Bien, "A dynamic gesture recognition system for the korean sign language (ksl)," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 26, no. 2, pp. 354–359, 1996.

9) Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," arXiv preprint arXiv:1311.2524, 2013.

10) Ronan Collobert and Jason Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in ICML. ACM, 2008, pp. 160–167.

11) Clement Farabet, Camille Couprie, Laurent Najman, ´ and Yann LeCun, "Learning hierarchical features for scene labeling," IEEE TPAMI, vol. 35, no. 8, pp. 1915– 1929, 2013. [15] Srinivas C Turaga, Joseph F Murray, Viren Jain, Fabian Roth, Moritz Helmstaedter, Kevin Briggman, Winfried Denk, and H Sebastian Seung, "Convolutional networks can learn to generate affinity graphs for image

segmentation," Neural Computation, vol. 22, no. 2, pp. 511–538, 2010.

12) Ao Tang, Ke Lu, Yufei Wang, Jie Huang, and Houqiang Li, "A real-time hand posture recognition system using deep neural networks," ACM Transactions on Intelligent Systems and Technology, 2014.