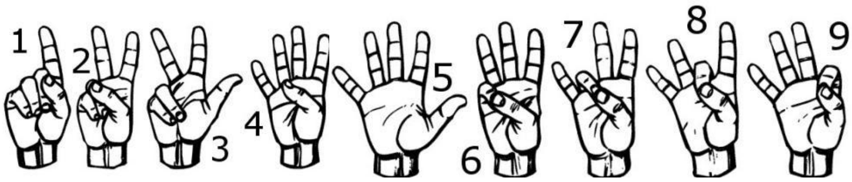
SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEUTRAL NETWORKS (CNN)

N.Koteswara Rao, Associate Professor , Department of CSE, Narayana Engineering College (Autonomous) Gudur,.SPSR Nellore, AP, India
K.Karthik Kumar Reddy , **G.Durgesh , K.Joyal Victor ,** UG Student, Department of CSE, Narayana Engineering College (Autonomous) Gudur,.SPSR Nellore, AP,

Abstract
Conversing to a person with aurally perceiving incapacitation is always a major challenge. Dactylology has indelibly become the ultimate panacea and is a very potent implement for individuals with aurally perceiving and verbalization incapacitation to communicate their feelings and opinions to the world. It makes the integration process between them and others smooth and less intricate. However, the invention of dactylology alone, is not enough . There are many strings annexed to this boon. The sign gestures often get mixed and discombobulated for someone who has never learnt it or kens it in a different language. However, this communication gap which has subsisted for years can now be narrowed with the prelude of sundry techniques to automate the detection of designation gestures . In this paper, we introduce a Dactylology apperception utilizing American Dactylology. In this study, the utilizer must be able to capture images of the hand gesture utilizing web camera and the system shall prognosticate and exhibit the designation of the captured image. We utilize the HSV colour algorithm to detect the hand gesture and set the background to ebony. The images undergo a series of processing steps which include sundry Computer vision techniques such as the conversion to grayscale, dilation and mask operation. The features extracted are the binary pixels of the images. We make utilization of Convolutional Neural Network(CNN) for training and to relegate the images. We are able to recognise 10 American Sign gesture alphabets with high precision. Our model has achieved a remarkable precision of above
90%.
Keywords: Sign Language, Hearing disability, Convolutional Neural Network(CNN), Web Camera, Machine Learning, Gesture recognition, Sign language recognition.

Introduction
As well stipulated by Nelson Mandela, “Talk to a man in a language he understands, that peregrinates to his head. Verbalize with him in his own language, that peregrinates to his heart”, language is indubitably essential to human interaction and has subsisted since human civilization commenced. It is a medium humans use to communicate to express themselves and understand notions of the genuine world. Without it, no books, no cell phones and definitely not any word I am inditing would have any construal. It is so deeply embedded in our everyday routine that we often take it for granted and don’t realize its consequentiality. Woefully, in the expeditious transmuting society we live in, people with auricularly discerning impairment are conventionally forgotten and left out[1][2]. They have to struggle to bring up their conceptions, voice out Dactylology, albeit being a medium of communication to auditorily impaired people, still have no denotement when conveyed to a non-dactylology utilizer. Hence, broadening the communication gap. To obviate this from transpiring, we are putting forward a designation language apperception system. It will be an ultimate implement for people with auricularly discerning incapacitation to communicate their cerebrations as well as a very good interpretation for non dactylology utilizer to understand what the latter is verbally expressing[3]. Many countries have their own standard and interpretation of denotement gestures. For instance, an alphabet in Korean dactylology will not designate an identically tantamount denotement to in Indian dactylology. While this highlights diversity, it withal pinpoints the involution of designation languages. American Dactylology is utilized to engender our datasets[4]. Figure 1 shows the American Sign Language alphabets. Identification of denotement gesture is perform.



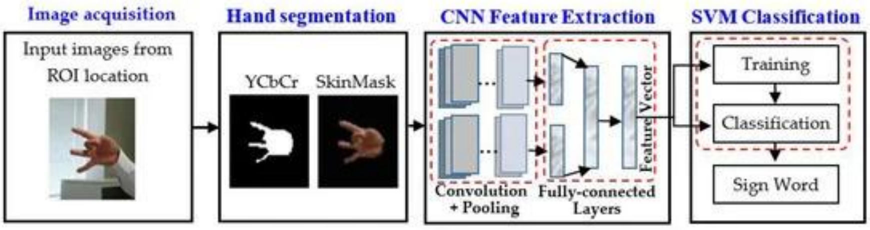
Numerical numbers in sign language

Existing system

Siming He proposed a system having a dataset of 40 prevalent words and 10,000 dactylology images. To locate the hand regions in the video frame, More expeditious R-CNN with an embedded RPN module is utilized. It amends performance in terms of precision. Detection and template relegation can be done at a higher speed as compared to single stage target detection algorithm such as YOLO. The detection precision of More expeditious R-CNN in the paper increases from 89.0% to 91.7% as compared to Expeditious-RCNN[5]. A 3D CNN is utilized for feature extraction and a designation-language apperception framework consisting of long and short time recollection (LSTM) coding and decoding network are built for the language image sequences.On the quandary of RGB dactylology image or video apperception in practical quandaries , the paper merges the hand locating network, 3D CNN feature extraction network and LSTM. This paper has achieved an apperception of 99% in prevalence lexicon dataset

Methodology

The first step of the proposed system is to amass data.Many researchers have utilized sensors or cameras to capture the hand forms of kineticism. For our system, we make utilization of the web camera to shoot the hand gestures.The images undergo a series of processing operations whereby the backgrounds are detected and eliminated utilizing the colour extraction algorithm Hue,Saturation,Value. Utilizing the morphological operations, a mask is applied on the images and a series of dilation and erosion utilizing elliptical kernel are executed . With openCV, the images obtained are amended to the same size so there is no distinction between images of different gestures . Our dataset has 2000 American sign gesture images out of which 2000 images are for training and the rest 500 are for testing purposes. Binary pixels are extracted from each frame, and Convolutional Neural Network is applied for training and relegation[6]. The model is then evaluated and the system would then be able to prognosticate the alphabets.

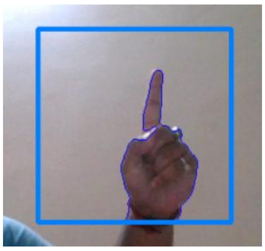


Data collection

Data accumulation is indelibly an essential part here as our result highly depends on it[7][8]. We have consequently engendered our own dataset having 2000 images of 10 static alphabet signs. We have 10 classes of static alphabets which are 1,2,3,4,5,6,7,8 and 9. Two datasets have been made by two different signs. Each of them performed one numerical gesture 500 times in alternate conditions. The dataset folder of alphabetic sign gestures is further split into 2 more folders, one for training datasets and the other for comparison. Out of the 2500 images captured, 2000 images are utilized for training and the rest for testing. To improve consistency, we have captured the pics in the same background with a camera each time a command is given. The images obtained are preserved in the png format .It is to be highlighted that there is no loss in quality whenever an pic in png format is opened ,closed and stored repeatedly.PNG is additionally good in handling high thresold and clear image. The camera will capture the images in the RGB colourspace.

Data processing

a.Image acquisition: Since the images obtained are in RGB colourspaces, it becomes more arduous to segment the hand gesture predicated on the skin colour only. We consequently transform the images in HSV colourspace. It is a model which splits the colour of an image into four separate components namely: Hue,Saturation and value. HSV is a potent implement to ameliorate stability of the images by setting apart effulgence from the chromaticity. The Hue element is impervious to any kind of illumination, shadows and shadings and can thus be considered for background abstraction. The region of the hand gesture undergoes dilation and erosionoperations with elliptical kernel.



b.Segementation: The first image is then transformed to grayscale. As much as this process will result in the loss of colour in the region of the skin gesture, it will withal enhance the robustness of our system to transmutations in lighting or illumination. Non-ebony pixels in the transformed image are binarised while the others remain unchanged, consequently ebony. The hand gesture is segmented firstly by taking out all the joined components in the image and secondly by letting only the component which is immensely connected, in our case is the hand gesture. At the terminus of the segmentation process, binary images of size 64 by 64 are obtained where the area in white represents the hand gesture, and the ebony coloured area is the rest.

c.Feature extraction: One of the most crucial part in image processing is to cull and extract consequential features from an image. Images when captured and stored as a dataset customarily take up a whole lot of space as they are comprised of an astronomically immense magnitude of data. Feature extraction avails us solve this quandary by abbreviating the data after having extracted the consequential features automatically. It additionally contributes in maintaining the precision of the classifier and simplifies its intricacy. In our case, the features found to be crucial, which are the binary pixels of the images. Scaling the images to 64 pixels has led us to get adequate features to efficaciously relegate the American Dactylology gestures . In total, we have 5096 number of features, obtained after multiplying 64 by 64 pixels.

Result:

Conclusion

Many breakthroughs have been made in the field of artificial astuteness, machine learning and computer vision. Many researches have been conducted on sign gesture apperception utilizing different techniques like ANN, LSTM and 3D CNN. On the other hand, our research paper requires low computing power and gives a remarkable precision of above 90%. In our research, we proposed to normalise and rescale our images to 64 pixels in order to extract features (binary pixels) and make the system more robust. We utilize CNN to relegate the 10 alphabetical American sign gestures and prosperously achieve a precision of 98% which is better than other cognate work verbalized in this paper.

Future work

We look forward to utilize more alphabets in our datasets and ameliorate the model so that it recognises more alphabetical features while at the same time get a high precision. We would withal relish to enhance the system by integrating verbalization apperception so that blind people can benefit as well.

References

[1] J. C. Cha, J. H. Cheon, “An identity-based signature from Diffie- Hellman groups,” PKC 2003, LNCS 25s7, 2003, pp.18–30.

[2] M. Bellare, J. A. Garay, T. Rabin, “Fast batch verification for modular exponentiation and digital signatures,” EUROCRYPT 1998, LNCS 1403, 1998, pp. 236–250.

[3] R. Gennaro, S. Jarecki, H. Krawczyk, T. Rabin, “Secure distributed key generation for discrete-log based cryptosystems,” EUROCRYPT 1999, LNCS 1592, 1999, pp.295–310.

[4] R. Guo, H. Shi, Q. Zhao, D Zheng, “Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems,” IEEE Access, PP 99, 2018, pp.11676–11686.

[5] Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. International journal of computer vision, 104(2), 154–171.

[6] Sucharita, V., Venkateswara Rao, P., Bhattacharyya, D., Kim, T.-H. Classification of penaeid prawn species using radial basis probabilistic neural networks and support vector machines International Journal of Bio-Science and Bio-Technology, 2016, 8(1), pp. 255–262

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[8] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431–3440).

[9] A.M.Mahaboob Basha, M.Rajaiah, Penchalaiah P et al., “Machine Learning-Structural Equation Modeling Algorithm: The Moderating role of Loyalty on Customer Retention towards Online Shopping”, International Journal of Emerging Trends in Engineering Research, ISSN: 23473983 (SCOPUS), Volume-8, No. 5, Page 1578–1585,May, 2020.