IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 11, No. 2, June 2022, pp. 525~529
ISSN: 2252-8938, DOI: 10.11591/ijai.v11.i2.pp525-529

Convolution neural networks for hand gesture recognation

Umesha Somanatti, Basavaraj A. Patil, Lingaraj Hadimani
Department of Computer Sciencel and Engineering, KLE Dr. M.S. Sheshgiri College of Engineering and Technology, Belagavi, India

Article Info

Article history:

Received Dec 22, 2020
Revised Feb 12, 2022
Accepted Mar 3, 2022

Keywords:

Convolution neural network
Finger segmentation
Hand gesture recognition
Human computer interaction

ABSTRACT

Hand gestures (not static or fixed positions) are movements of fingers and the arm to communicate messages. Hand gesture recognition is the process of identifying meaningful expressions involving the human hand. Pictorial representation of gestures will enable to understand human computer interaction (HCI). This paper describes a system using convolution neural network (CNN) for recognizing the 26 letters of the English alphabet signaled with hand gestures. A Python program was developed to recognize the gestures made in front of a web camera. The hand gestures obtained are categorized using CNN with a trained model. The model was constructed using 1,100 gestures images. The recognition rate was obtained with 91% of accuracy. The proposed method was found to be highly efficient in distinguishing and classifying gestures.

Corresponding Author:

Umesha Somanatti
Department of Computer Science and Engineering, KLE Dr. M.S. Sheshgiri College of Engineering and Technology
Udyambag Belgaum 590008, India
Email: usomanatti@gmail.com

1. INTRODUCTION

Recent advances in information technology and computer systems have deeply impacted our day-to-day life. One recent application of information technology, which has great potential, is in the interaction between human and computer. Gesture recognition is a natural communication tool, offering a powerful means of interaction humans and computers. Traditional means of input such as keyboards and mouse reduce the speed of communication between computer and human. On the other hand, hand gestures can be used to recognize the letters of the English alphabet.

Hand gestures are an indispensable means of communication for people who are speech and hearing impaired. In computers, recognition of continuous gesture patterns is possible by using an artificial neural network (ANN) [1]–[3]. One advantage of using hand gestures in computers is that visual interpretation will help in user ease and spontaneity in human computer interaction (HCI) [4]–[7]. This study describes an accurate gesture detection system designed for use with convolution neural network (CNN). Possible applications of this system include computer games, machinery control and related uses. Proposed work does not need gloves with special sensors. We use video graphic array (VGA) camera to capture the hand gestures. To acquire data some, hand gesture recognition systems need data glove [8], [9]. In a gesture recognition system, the motion of a person's fingers or arms are used to convey information. A hand gesture recognition system interprets the meanings conveyed by gestures. In such make it systems, the features of gestures are extracted from images and are used to form feature vectors. These vectors are mapped to the original data set. There are several gesture recognizing software among which the most common is the American sign language (ASL) [10], [11]. ANN are parallel distributed processors with simple processing units called neurons. ANN is used to acquire, store and utilize the knowledge which is acquired through learning or training, which helps ANN to train itself for all possible cases. ANN is widely used in applications like image and voice recognition [12], [13].

2. RELATED WORK

Maung developed a supervised neural network system using MATLAB toolbox to recognize real time gestures of the Myanmarese alphabet. The system was designed for speed and did not use complex hardware. The input images for the system were digitized photographs with which feature vectors were generated using histograms. The vectors in turn were fed into the neural network (NN) system. Because of the MATLAB tool, although the design was not complex, the time taken for implementation was large [14].

In another work, Chen *et al.* presented a new method for hand gesture recognition wherein a background subtraction method was utilized to detect the hand region from the background. Further, by using a segmentation technique, fingers and palms were segmented. A simple rule-based classifier was used to recognize the hand gestures. The proposed algorithm, yielded a good overall accuracy of 96.6% on the dataset of 1,300 images [15].

A paper presented by Fu *et al*. described a wavelet-based image preprocessing technique for gesture recognition. The authors demonstrated a method for feature extraction, which was tested with six different hand gestures. Their paper described methods for obtaining 1-dimensional signals using 2-dimensional hand gesture contour images. For 1-dimensional signals, the system decomposes the wavelets. The system could also extract statistical features of the wavelet coefficients. However, the conversion to 1-D conversion from 2-D affected the accuracy of neural network and thus, could be applied only to a few hand gestures [16].

Yamato *et al.*, in their paper, discussed a system that could recognize gestures using three models. The results obtained from each model are integrated to obtain a composite result. In this model, audio and motion are learned by the hidden markov model (HMM), whereas random forest (RF) is used to learn the video model. Here the uni-modal and multi-modal models ware compared for determining the accuracy of recognition [17].

Bobic *et al.* proposed a method of recognition of hand gestures using neural networks. The authors used multiple background and space orientations to capture images. A histogram of oriented gradients was used for feature extraction and backpropagation algorithm for training. In another method, the authors implemented a sparse auto encoder. In this method, more gestures were used for training and less for recognition. Another limitation was that the authors static hand gestures in their study [18].

Badi *et al.,* in their paper, proposed that images are pre-processed and classified using ANN. During the preprocessing stage, edge detection, homogeneity, and other filtering operations ware performed. And then by using complex hand contour and Ahzat methods the lines are extracted from the hand gestures [19].

3. PROPOSED SYSTEM

The hand gesture recognition experiment was conducted using web camera and required a white background with sufficient illumination. A particular gesture was shown in front of camera and action was identified by the trained model. This process had both training and testing phase. In Figure 1, we shown the system diagram of proposed work. Different hand gestures were captured and were provided to the system as an input. The proposed method used desktop system and web camera. To obtain the gestures, user had to show his hand in front of camera. From video frame red green blue (RGB) image was extracted. Then these images were converted into hue, saturation and value (HSV) type.

The major applications of CNNs are for image recognition, pattern recognition, speech recognition and natural language problems [20]. A Convolution neural model consists of one or more convolution layers, pooling layers and fully connected layers. Kernel convolution is used in CNNs. It is the process where we take a small matrix of numbers and we call it as filter or kernel. Then it is passed over an image and transforms it. The main objective of the proposed work was to design a system for recognition of hand gestures for 26 English alphabets using CNN with equitable accuracy. Subsequent feature map (f) values are calculated according to following formula where the input image is denoted by X and our kernel by h:

$$f[a,b] = (X * h)[a,b] = \sum_{m} \sum_{n} X[a-m, b-n] \times h[m,n] \tag{1}$$

where m, n are matrix indices.

The CNN used in our system has seven layers-three convolutional, three max-pooling layers, and one fully connected. Our first convolutional layer consists of 32 filters of size 3×3. We used rectified linear unit (ReLU) activation. ReLU function is a mathematical calculation which clips negative values to zero and if it returns positive values will be unchanged. Mathematically, it is represented as

$$\mathrm{Relu}(x) = max(0, x) \tag{2}$$

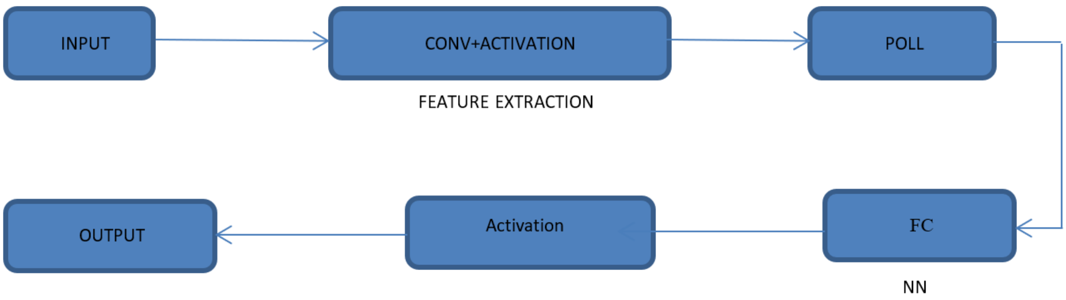


Figure 1. System diagram of hand gesture recognition

The next layer is max polling layer. Max polling layer provides the reduction in spatial dimension (length and width). It is used to reduce size of image by taking the maximum value in the window. In our system a max pooling layer of 2×2 was used with a stride of two in both directions. Thirty-two filters of size 3×3 and ReLU activation were used in the second convolutional layer of our system. The pool size of a second max pooling layer was 2×2. Sixty-four filters of size 3×3 comprise the third convolutional layer.

The max pooling layer used is with pool size 2×2. Max pooling layers help to reduce the number of parameters for large images. Max pooling take the largest element from the feature map. The fully connected layer is the final one, where all the neurons of present layer and the neurons of next layer are connected with each other. The fully connected layer feed forward neural network helps in computing class scores. Input to this layer is from the last pooling layer. The output of the pooling layer is flattened and applied to feed forward neural network. Flattening is the process of unrolling matrix values into vectors. At every layer, the following calculation takes place:

$$o = \sum_{i=1}^{m} Wi \times Xi + b \tag{3}$$

where $\boldsymbol{Xi}$ is input vector, $\boldsymbol{Wi}$ is weight vector and $\boldsymbol{b}$ is bias.

This is followed by SoftMax activation. The Soft-max function used here is for classification purpose. Probability distribution is the output soft-max layer that is the values of the output sum to 1. The output of softmax function represents a distribution over class labels. It also obtains the probability of each input element belonging to a label. Mathematical model of SoftMax is

$$\sigma(z)j = \frac{e^{zj}}{\sum_{1}^{k} e^{zk}} \text{ for } j = 1\ldots.k \tag{4}$$

Gesture recognition problem has 26 neurons in the output layer that means we have 26 classes.

4. RESULTS

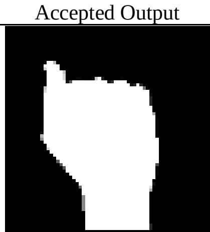
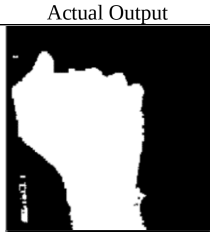
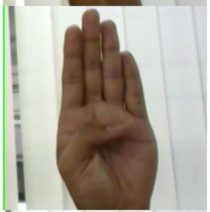
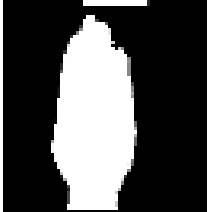
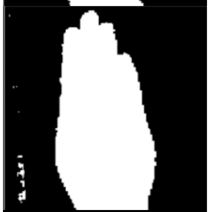
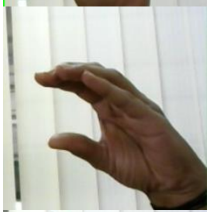
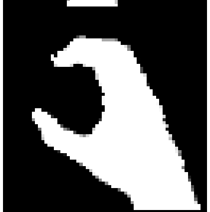
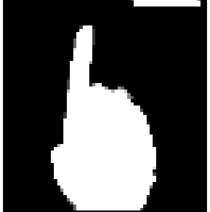
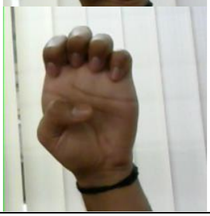
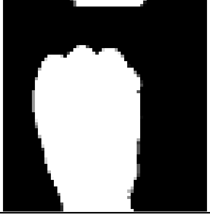
We implemented convolution neural network using keras library and Google Tensor Flow [21]–[25]. The gesture recognition system was trained with the sign language provided in Modified National Institute of Standards and Technology (MNIST) dataset. Six rounds of experiments were carried out using a different training and testing mode each time to obtain optimum results. The test results obtained from our experiments are presented in Table 1 (see in appendix). Training accuracy obtained was 91%.

5. CONCLUSION

We have proposed and tested a web camera-based approach for hand gesture recognition using convolution neural network to recognize different hand gestures. Our model was evaluated using a hand gesture dataset. The results of our experiments demonstrate that gesture recognition can attain 91% accuracy.

APPENDIX

Table 1. Test results

Test Case ID	Description	Input	Accepted Output	Actual Output
A	This is the 1st letter from ASL in the dataset.			
B	This is the 2nd letter from ASL in the dataset.			
C	This is the 3rd letter from ASL in the dataset.			
D	This is the 4th letter from ASL in the dataset.			
F	This is the 5th letter from ASL in the dataset.			

REFERENCES

[1] S. Mitra and T. Acharya, "Gesture recognition: a survey," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, May 2007, doi: 10.1109/TSMCC.2007.893280.

[2] M. R. Malgireddy, J. J. Corso, S. Setlur, V. Govindaraju, and D. Mandalapu, "A framework for hand gesture recognition and spotting using sub-gesture modeling," in *2010 20th International Conference on Pattern Recognition*, Aug. 2010, pp. 3780–3783, doi: 10.1109/ICPR.2010.921.

[3] P. Suryanarayan, A. Subramanian, and D. Mandalapu, "Dynamic hand pose recognition using depth data," in *2010 20th International Conference on Pattern Recognition*, Aug. 2010, pp. 3105–3108, doi: 10.1109/ICPR.2010.760.

[4] C. Keskin, F. Kıraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," in *Consumer Depth Cameras for Computer Vision*, London: Springer London, 2013, pp. 119–137.

[5] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1110–1120, Aug. 2013, doi: 10.1109/TMM.2013.2246148.

[6] A. M. H. Wong and D.-K. Kang, "Stationary hand gesture authentication using edit distance on finger pointing direction interval," *Scientific Programming*, vol. 2016, pp. 1–15, 2016, doi: 10.1155/2016/7427980.

[7] K. N. Shah, K. R. Rathod, and S. J. Agravat, "A survey on human computer interaction mechanism using finger tracking," *International Journal of Computer Trends and Technology*, vol. 7, no. 3, pp. 174–177, Jan. 2014, doi: 10.14445/22312803/IJCTT-V7P148.

[8] Y. Wang, C. Yang, X. Wu, S. Xu, and H. Li, "Kinect based dynamic hand gesture recognition algorithm research," in *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Aug. 2012, pp. 274–279, doi: 10.1109/IHMSC.2012.76.

[9] G. Dewaele, F. Devernay, and R. Horaud, "Hand motion from 3D point trajectories and a smooth surface model," in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2004, pp. 495–507.

[10] V. Frati and D. Prattichizzo, "Using Kinect for hand tracking and rendering in wearable haptics," in *2011 IEEE World Haptics Conference*, Jun. 2011, pp. 317–321, doi: 10.1109/WHC.2011.5945505.

[11] Z. Meng, J.-S. Pan, K.-K. Tseng, and W. Zheng, "Dominant points based hand finger counting for recognition under skin color extraction in hand gesture control system," in *2012 Sixth International Conference on Genetic and Evolutionary Computing*, Aug. 2012, pp. 364–367, doi: 10.1109/ICGEC.2012.85.
[12] K. Hu, S. Canavan, and L. Yin, "Hand pointing estimation for human computer interaction based on two orthogonal-views," in *2010 20th International Conference on Pattern Recognition*, Aug. 2010, pp. 3760–3763, doi: 10.1109/ICPR.2010.916.
[13] A. D. Bagdanov, A. Del Bimbo, L. Seidenari, and L. Usai, "Real-time hand status recognition from RGB-D imagery," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 2456–2459.
[14] T. H. H. Maung, "Real-time hand tracking and gesture recognition system using neural networks," *World Academy of Science, Engineering and Technology*, pp. 466–470, 2009.
[15] Z. Chen, J.-T. Kim, J. Liang, J. Zhang, and Y.-B. Yuan, "Real-time hand gesture recognition using finger segmentation," *The Scientific World Journal*, vol. 2014, pp. 1–9, 2014, doi: 10.1155/2014/267872.
[16] X. Fu, J. Lu, T. Zhang, C. Bonair, and M. L. Coats, "Wavelet enhanced image preprocessing and neural networks for hand gesture recognition," in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, Dec. 2015, pp. 838–843, doi: 10.1109/SmartCity.2015.172.
[17] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 379–385, doi: 10.1109/CVPR.1992.223161.
[18] V. Bobic, P. Tadic, and G. Kvascev, "Hand gesture recognition using neural network based techniques," in *2016 13th Symposium on Neural Networks and Applications (NEUREL)*, Nov. 2016, pp. 1–4, doi: 10.1109/NEUREL.2016.7800104.
[19] H. Badi, A. Hamza, and S. Hasan, "New method for optimization of static hand gesture recognition," in *2017 Intelligent Systems Conference (IntelliSys)*, Sep. 2017, pp. 542–544, doi: 10.1109/IntelliSys.2017.8324347.
[20] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, no. 2, pp. 743–755, Apr. 2020, doi: 10.1007/s11036-019-01445-x.
[21] N. Ketkar, "Introduction to keras," in *Deep Learning with Python*, Berkeley, CA: Apress, 2017, pp. 97–111.
[22] B. B. Traore, B. Kamsu-Foguem, and F. Tangara, "Deep convolution neural network for image recognition," *Ecological Informatics*, vol. 48, pp. 257–268, Nov. 2018, doi: 10.1016/j.ecoinf.2018.10.002.
[23] S. Preetha Lakshmi, S. Aparna, V. Gokila, and P. Rajalakshmi, "Hand gesture recognition using CNN," in *Smart Innovation, Systems and Technologies*, Springer Singapore, 2022, pp. 371–382.
[24] K. Ramasubramanian and A. Singh, "Deep learning using keras and TensorFlow," in *Machine Learning Using R*, Berkeley, CA: Apress, 2019, pp. 667–688.
[25] F. J. J. Joseph, S. Nonsiri, and A. Monsakul, "Keras and TensorFlow: a hands-on experience," in *Advanced Deep Learning for Engineers and Scientists*, Springer International Publishing, 2021, pp. 85–111.

BIOGRAPHIES OF AUTHORS

Umesha Somanatti M.E CSE, Associate Professor, with 23 years of teaching and research expierence. Has attended and persented research papers in conferences. Guided several UG and PG students in their research project. He can be contacted at email: usomanatti@gmail.com.

Basavaraj A. Patil M.E CSE, Associate Professor, Working r in CSE department at KLE Dr. MSSCET. 25 years of teaching experience. He attended and persented research papers in in many conferences. Guided several students in their research project. He can be contacted at email: b_a_patil@yahoo.com.

Lingaraj Hadimani M Tech, Ph.D CSE, Associate Professor, Working in CSE department at KLE Dr. MSSCET. 15 years of teaching experience. He attended and persented research papers in renowned conferences and published papers in reputed journals. Guided several students in their research project. He can be contacted at email: lahadimani@gmail.com.