International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Transitioning from Convex Hull Algorithm to CNN

Tanmesh Mishra[1], Suhasini S Rao[1], Thanmayi M P[1], Triveni Bhat[1], Shwetha S[2]

[1]*Student, Dept. of Information Science Engineering, NIE, Mysuru, Karnataka, India*
[2]*Assistant Professor, Dept. of Information Science Engineering, NIE, Mysuru, Karnataka, India*

Abstract – *The paper aims to compare the performance of the proposed system using two algorithms, Convex Hull and CNN. The fabricated model recognizes the gestures shown by the hearing-impaired and returns the recognized alphabet or number. The gestures must belong to the American Sign Language (ASL) for the model to acknowledge it. The goal is to encompass the transition in the selection of the appropriate algorithm for the superior functioning of the model.*

Key Words: Convex Hull, CNN, ASL, Hand Gesture, Epoch

1. INTRODUCTION

Hands are used most frequently by human beings to communicate and interact with machines. Mouse and Keyboard are the basic input/output to computers and the use of both of these devices require the use of hands. Although hands are most commonly used for day to day physical manipulation related tasks, but in some cases, they are also used for communication. Hands are most important for the hearing-impaired, who depend on their hands and gestures to communicate, so hand gestures are vital for communication in Sign Language. If computer had the ability to translate and understand hand gestures, it would be a leap forward in the field of human computer interaction. The aim is to create a model that recognizes the gestures and produce the results with maximum accuracy and efficiency. The model uses two algorithms independently and then compares the performance of the models.

2. PROBLEM STATEMENT

A utilizable tool to recognize hand gestures is essential as only a small portion of the population is aware of it. In an effort to make the life of the hearing-impaired easier, a model that recognizes and translates the gestures is required. The model should also provide a user friendly interface and must provide optimal performance with high accuracy.

3. LITERATURE SURVEY

1. Yuqian Chen and Wenhui Zhang," Research and implementation of sign language recognition method based on Kinect," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2016, pp. 1947-1951.The system proposed in this paper uses a Kinect camera. Kinect is a 3D somatosen-sony camera launched by Microsoft, which can capture the color, depth and skeleton frames and is helpful to the gesture recognition research. It is shown in the experimental results that the use of the HOG and SVM algorithms significantly increases the recognition accuracy of the Kinect and is insensitive to background and other factors. The method realized in this paper has a great significance for sign language recognition technology.
2. K. K. Dutta, Satheesh Kumar Raju K, Anil Kumar G S and Sunny Arokia Swamy B," Double handed Indian Sign Language to speech and text," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 374-377. The proposed system in this paper proposes a system that provides voice to the deaf and mute people and promising them an independent life without any help of human translator. The system is trained with double handed sign language by using a mini-mum eigenvalue algorithm. Here Logitech web camera is used for image acquisition and processing is performed in MATLAB. The corresponding output is obtained after extracting Shi-Thomasi corner feature detector

4. GESTURE RECOGNITION USING CONVEX HULL

The proposed system is a device that has a camera that captures the hand signs made by the person whose actions are processed and recognizes it to text and then person who is hearing impaired. Every gesture has some distinct features, which differentiates it from other gestures, moments are used to extract these features of gestures and then classify them using CONVEX HULL algorithm.

4.1 Execution

The execution is done using tools OpenCV in Python, Python Tkinter and PyCharm. The real time image will be captured from the camera. Then a RGB to grey conversion is performed on the captured image. The grey image is smoothened by softening in sharp edges which is followed by thresholding process where a division of images is done. A contour finding method is implemented to find the extract the boundaries. Then convex defects are determined using convex hull algorithm to determine the complete characteristics of the captured image in digital form and thus completing the processing to be applied on the image.

If the processed image and the templates of the gestures got matched, the corresponding gesture is recognized, and the corresponding text message will be obtained. Hence this setup can efficiently provide us aid to build the conversation with the hearing-impaired people.

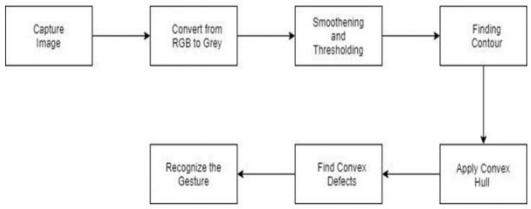

Fig -1: Block diagram of the proposed system using Image Processing

4.2 Advantages of Image Processing using Convex Hull

1. The recognition of the gesture and the processing of the capture images is done real time .As a result the user gets the result of the input in meantime without much delay.
2. The main purpose of Image Processing is removal of noise as it hinders the performance of the algorithm and does not provide any insights on the recognition of the hand gesture.
3. The resulting image after gray scaling needs lesser storage memory and also provides ease of access.
4. The method resulted in recognition of a partial set of the alphabets and numbers

4.3 Disadvantages of Image Processing using Convex Hull

1. The implementation of the method takes up more CPU as the processing has to be done for every new gesture shown by the user in the camera.
2. This is less efficient as the processing is redundant and slow.
3. The machine never learns about the gesture, it compares the predetermined value associated with the area of the hand gesture to determine the correct alphabet or number.
4. It does not work for all the alphabets and numbers as any two gesture can occupy the same amount of area in the vision of the computer.
5. Alphabets and numbers cannot be recognized together

5 GESTURE RECOGNITION USING CNN

The neural network mimics the structure of the neurons of the human brain. The network is built in layers of nodes which are interconnected with each other using activation functions. The classes of Neural network covers several architectures such as Convolutional Neural Networks(CNN), Recurrent Neural Networks (RNN). Therefore, CNN is just one kind of Neural network the first part of CNN is the convoluted part [9]. It functions as feature extractor of the image. Here, an image is passed through a succession of filters which extract the essential information from it along with reducing its resolution. It consists of three layers, CONV layer, RELU layer and the Maxpool layer

- CONV layer - The objective of a Conv layer is to extract features of the input volume This layer accepts input of size [W1xH1xD1], where W1 is the width, H1 is the height, and D1 is the depth, and the output is calculated by the product between the weights and the local region, they are connected to in the input layer. The obtained output is of size [W2xH2xD2] called as the 'convolution map' were W2, H2, and D2 are the width, height, and depth of the input size, respectively
- ReLU layer - This layer applies an element-wise activation function max(0,x), which turns negative values to zeros. This layer increases the nonlinear properties of the model and the overall network without affecting the receptive fields of the conv layer. This layer does not change the dimension
- Maxpool layer - This layer performs a function to reduce the spatial dimensions of the input [12], and the computational complexity of our model. Moreover, it also controls overfitting. If the input dimensions are [W1×H1×D1], and the output dimensions are [W2×H2×D2]
  $W2 = (W1 - F)/S + 1$ (4)
  $H2 = (H1 - F)/S + 1$ (5)
  $D2 = D1$ (6)

5.1 Execution

We have used the Sequential model to build layers in both the architecture, which is the easiest way to build a model in Keras [17]. We use the categorical crossentropy loss function for single label categorization and used 'SGD' optimizer or Stochastic gradient descent simply does away with the expectation in the update and computes the gradient of the parameters using only a single or a few training examples. The SGD optimizer includes support for momentum, learning rate decay, and Nesterov momentum. We have also used Batch size of 32 and epoch value is 25. The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters, and we have used 32 as Batch size as is a good starting point, we could have also tried with 64, 128, and 256. The number of epochs is a hyperparameter that defines the number times that the learning algorithm work through the entire training dataset. The number of epochs is traditionally large, often hundreds or thousands, allowing the learning algorithm to run until the error from the model has been sufficiently minimized.

The Convolutional Neural Network architecture uses ResNet50 architecture along with the Sequential model. The ResNet-50 is a convolutional neural network that is trained on more than a million images from the ImageNet database. The network is 50 layers deep and can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. Using this gives us the advantage of using pre- trained models as effective feature extractors for new images to solve diverse and complex computer vision tasks [18]. A way to save time in model training is to use this model and extract out all the features from our training and validation data-sets and then feed them as inputs to our classifier.

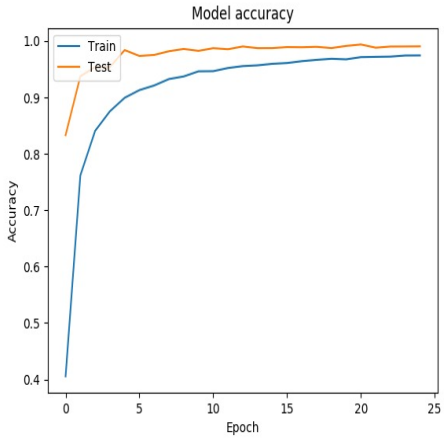

Fig –2 : Accuracy vs Epoch

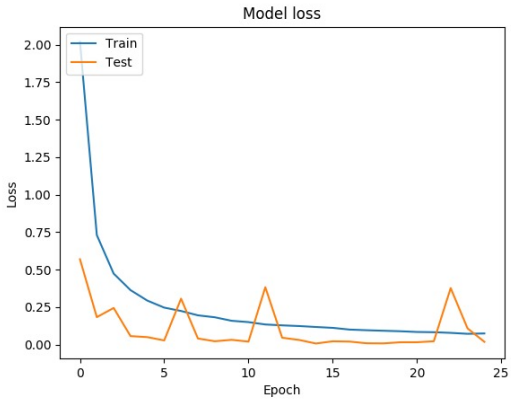

Fig -3: Loss vs Epoch

An epoch is a term used in machine learning and indicates the number of passes through the entire training dataset the machine learning algorithm has completed. The following figure shows how the model improved with increase in epoch.

epoch	accuracy	loss
0	0.025	3.4673723936080900
0	0.0940625	3.3543837761879000
0	0.0675	3.4101862859726000
1	0.1665625	3.0192066287994400
0	0.0909375	3.38825236082077
1	0.1775	2.9677642393112200
0	0.405625	2.0192474120110300
1	0.7617578	0.7299730169773100
2	0.84069586	0.47281053387811200
3	0.8753125	0.3626991576794540
4	0.89929634	0.2927311514298180
5	0.9129297	0.24606875360477700
6	0.9211493	0.22355727394686000
7	0.9325645	0.19457393425623800
8	0.9371875	0.18162717520957800
9	0.94625	0.15822858653962600
10	0.9465234	0.14935808679496400
11	0.9521501	0.13358125056870500
12	0.9555078	0.12774401092378000
13	0.9568804	0.12314580132117300
14	0.9595703	0.11682256140426000
15	0.9609851	0.11105020296608100
16	0.96425784	0.09951458752533650
17	0.96660155	0.09532438542024470
18	0.968491	0.0917596476398659
19	0.9675	0.08847372609408920
20	0.971423	0.08365233227220290
21	0.97191405	0.08206181316854780
22	0.97234374	0.07763987874655870
23	0.9743159	0.0713358084333513
24	0.9744331	0.07375262392352770

Fig -3: Improvement in model with increase in Epoch

6. CONCLUSION

The conclusion obtained is that with the help of CNN better performance of the model can be achieved. The model is trained using the sample datasets and hence it provides the results with more accuracy.

REFERENCES

[1] Redmon J, and Angelova A, "Real-time grasp detection using convolutional neural networks", IEEE International Conference on Robotics and Automation, pp. 1316–1322, 2015.

[2] H.B. Burke, P.H. Goodman, D.B. Rosen, D.E. Henson, J.N. Weinstein, F.E. Harrell, J.R. Marks, D.P. Winchester, D.G. Bostwick, "Artificial neural networks improve the accuracy of cancer survival prediction." Cancer, vol. 79, pp. 857-862, 1997.

[3] Hang Chang, Cheng Zhong, Ju Han, Jian-Hua Mao, "Unsupervised Transfer Learning via Multi-Scale Convolutional Sparse Coding for Biomedical Application." IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 janvier 2017.

[4] Howard, A., "Some improvements on deep convolutional neural network based image classication." ICLR, 2014.

[5] van de Sande, K. E. A., Gevers, T., and Snoek, C. G. M, "Evaluating color descriptors for object and scene recognition", IEEE Transactions on Pattern Analysisand Machine Intelligence." 1582– 1596. 2010.

[6] V. Singh and S. P. Lal, "Digit recognition using single layer neural network with principal component analysis," Asia-Pacific World Congress on Computer Science and Engineering, Nadi, 2014, pp. 1-7.

[7] G. Wang, G. B. Giannakis and J. Chen, "Learning ReLU Networks on Linearly Separable Data: Algorithm, Optimality, and Generalization," in IEEE Transactions on Signal Processing, vol. 67, no. 9, pp. 2357-2370, 1 May1, 2019.

[8] M. S. Imtiaz and K. A. Wahid, "Image enhancement and spacevariant color reproduction method for endoscopic images using adaptive sigmoid function," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, 2014,

[9] N. Jmour, S. Zayen and A. Abdelkrim, "Convolutional neural networks for image classification," 2018 International Conference on Advanced Systems and Electric Technologies (IC), Hammamet, 2018.

[10] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel. "Backpropagation applied to handwritten zip code recognition." Neural Computation, 1(4):541–551, 1989.

[11] Y.Boureau, F. Bach, Y. LeCun, andJ. Ponce. "Learning midlevel features for recognition." CVPR, 2010.

[12] H. B. Burke, "Artificial neural networks for cancer research: Outcome prediction." Sem. Surg. Oncol, vol. 10, pp. 73–79, 1994.

[13] A. Ahmed, K. Yu, W. Xu, Y. Gong, and E. Xing. "Training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks." ECCV,2008.

[14] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016.

[15] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L., "ImageNet: A large-scale hierarchical image database." In CVPR, 2009.

[16] Ahonen, T., Hadid, A., and Pietikinen, "M. Face description with local binary patterns: Application to face recognition." Pattern Analysis and Machine Intelligence, 2037–2041. 2016.

[17] S. Arora and M. P. S. Bhatia, "Handwriting recognition using Deep Learning in Keras," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida (UP), India, 2018, pp. 142-145.

[18] M. Shaha and M. Pawar,"Transfer Learning for Image Classification," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 656-660.

[19] Yuqian Chen and Wenhui Zhang," Research and implementation of sign language recognition method based on Kinect," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2016, pp. 1947-1951

[20] K. K. Dutta, Satheesh Kumar Raju K, Anil Kumar G S and Sunny Arokia Swamy B," Double handed Indian Sign Language to speech and text," 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, 2015, pp. 374-377.

[21] Umang Patel and Aarti G. Ambekar , "Moment Based Sign Language Recognition For Indian Languages," 2017 Third International Conference on Computing, Communication, Control And Automation (ICCUBEA)

[22] J. L. Rahejaa, A. Mishrab, and A. Chaudharyc," Indian Sign Language Recognition Using SVM," ISSN 10546618, Pattern Recognition and Image Analysis,2016, Vol. 26, No. 2, pp. 434441. Pleiades Publishing, Ltd., 2016.

[23] Shangeetha R.K,Valliammai V and Padmavathi S,"Computer Vision Based Approach For Indian Sign Language Character Recognition", IEEE 2012.

[24] Ali. A Abed and Sarah A Rahman," Python based Raspberry Pi for Hand Gesture Recognition". Research Gate ijca2017915285. International journal of Computer Applications. September 2017.

[25] Prof. R.R Itkarkar and Prof. Anil V Nandi. " Hand Gesture to Speech Conversion using MATLAB". IEEE - 31661. 4th ICCCNT 2013, July.

[26] 8. Rajat Shrivastava , A Hidden Markov Model based Dynamic Hand, IEEE 2012journal.

[27] Vincent Muehler's webpage on hand gesture recognition using OpenCV - https://medium.com/@muehler.v/simple-hand-gesture-recognit ion-using-opencv-and-javascript-eb3d6ced28a0