JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# A New Approach To Hand Gesture Recognition Using Combined Effect Of CNN And HMM

**Nandini M S[1], Salila Hegde[2], Ashadeepa[3]**
[1]Dept. of IS & Engg., NIE North campus, Mysuru, Karnataka, India,
[2]Dept. of EC & Engg., NIE North campus, Mysuru, Karnataka, India,
[3]Dept. of CS & Engg., VVCE, Mysuru, Karnataka, India,

**Abstract**—Gesture language is a communication language that involves hand gestures, body language and facial expression, used basically by the people who have hearing impairment. They communicate among themselves in sense of letters, words and sentences of interactive language to gestures. The most expressive way to communicate among themselves is Sign Language. Here, proposed system will recognize the hand gestures and translates into textual words. The methodology consists of two phases, namely Model Creation phase a Recognition phase. Here, Convolution Neural Network(CNN) is used for building the model and Hidden Markov Model(HMM).
**Index terms —Sign language, hearing impairment, hand gestures, CNN, HMM.**

## I. INTRODUCTION

Gestures play a vital role in daily life of differently-abled people and it helps to convey information and express people feelings. Hand gesture tracking and identification is an vital area of research in human computer interaction .Hearing impaired people becomes isolated as normal people fail to interact with them due to their ignorance of sign language. This will create bad impact on their social and working life. A translator is badly needed when a person wants to interact with hearing impaired person, but there is a short fall of such experienced and educated interpreter. Hence, there is a need of automated system that can interpret the sign language and help differently-abled in interacting with normal people. Gestures may be categorized as dynamic and static[1]. A dynamic gesture tends to change over a periodof time and static gesture tends to sustain almost unchanged over time. Here, the proposed system focuses on recognizing the static gestures. The main problem is how to make a machine able to track the hand movements. Hand gestures differentiates in position of fingers and hand movements. So, we need to focus on non- linearity as it is one of the qualities of hand gestures. The metadata of hand movement images are used to identify the gestures. A CNN runs by getting features from images. This overcomes the need for manual way accessing feature. The features are not trained features. They are trained during the network trains on a group of images. This made deep learning models extremely accurate for computer vision tasks. Hence we have used CNN in proposed system. Along with CNN, HMM also been employed in our proposed system. HMM is used topredict hidden state. It allows us to detects a sequence of hidden variables from a set of observed variables.

### 1.1 RELATED WORK

Many research works have been done on hand gesture identification.

Glove based system [2, 9] have been developed for recognizing sign languages. In this system the user has to wear a device which carries a load so it establishes connection with the device to a computer. Such devices are costlier and decreases the naturality of the gesture language communication. System has been developed using multi-class Support Vector Machine (MSVM) [8, 12] for training and recognizing signs of ISL. It works on three phases, namely a training , a testing and a recognition. Combined parameters of Hu invariant moment and structural shape descriptors are used to make a new feature vector to identify hand movement.

A gesture recognition system using ANN based on shape fitting technique has been developed [4,16]. In this system, a segmentation technique on color space was used after filtering to detect hand. Then the movement of the hand was approached by the hand analysis. The movements of hands and finger inclination features were extracted and moved to an ANN. System developed using Kinect [3,4], here recognition of isolated Polish Sign Language words detected by Kinect. A whole word model approach with KNN classifier applying DTW technique is differentiated against the methodology using models of subunits..

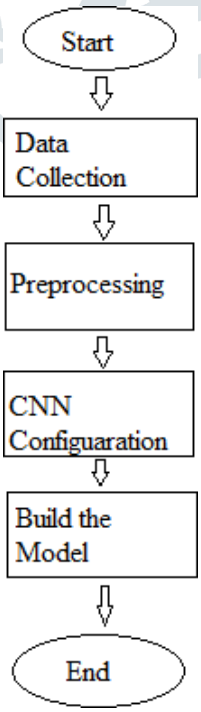
Hand movements detection by using an ANN has built [5,17]. In this system, images were portioned based on skin colors. The identified features for ANN were change in pixel through cross sections, boundary and the scalar description. After getting those feature vectors, they were inputted to the ANN for learning. An approach based on haar-like features was proposed [6]. In this system, AdaBoost algorithm was used to learn the model. The complete process was divided into two stages.

However, there are few drawbacks in manual feature extraction. The extraction process is complex and not all the possible feature are extracted. The extraction might become human-biased. Further, automated feature engineering systems emerged which are simple and not human biased . Also, almost each and every feature can be captured using automat system. Features from structured data can be extracted using CNN. So, change-over to automated feature engineered was made.

## 2. Approached technique

The design of system is divided into two phase. One is Model Creation Phase and second is Recognition Phase. These two phases are shown in Figures 1 and 2 respectively. The approach is the hybrid effect of data collection, pre-processing, configuring the CNN , developing the model and recognizing the hand gesture and Predicting the meaning of sequence of inputs using Hybrid CNN-HMM .

**Fig. 1.** Model Creation Phase.



## 3. DATASET AND TRAINING

Its a self created dataset for Indian sign language. It consists of 1575 training images and 175 testing images that belonging to 9 classes. Each class represents a static gestures to recognize. Here, collected images are needed to train and validate the model. The gestures were collected from different individuals. The input images has single hand, gesture were shown with right hand or left hand, the palm facing the camera. The recognition process will be easier if there is less complexity in image background and images were of high resolution with high contrast on the hand.

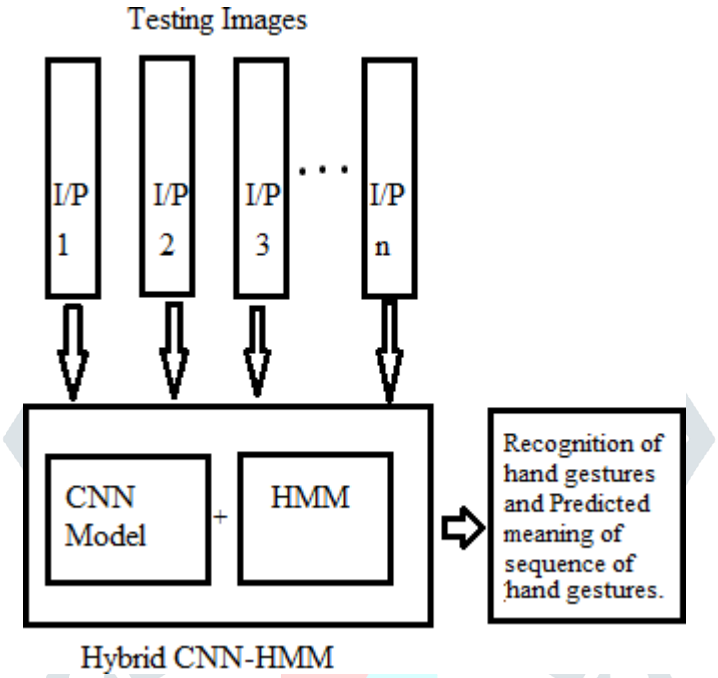



**Fig. 2**. Recognition Phase

## 4. PRE-PROCESSING

Dataset undergoes pre-processing to minimize the computational complexity and to achieve increased performance rate. At first, background of the images were deleted by the technique of background subtraction built in [8,9]. The subtraction of background is based on the technique K-Gaussian distribution. After background subtraction, all that remains is the images of hand. Then the images were transformed to grayscale. Since grayscale images contain only one color channel, it will be easier for CNN to learn [10]. Then it undergoes morphological erosion[12]. Noise has reduced using filtering process. While processing the signal, its often needed to reduce noise [11]. The images were then resized to 256x256 and then fed to CNN.

## 5. CNN CONFIGURATION

The CNN is adopted in this research work to recognize the hand gesture. It consists of four convolution layers, four max pooling layers, two fully connected layers and a output layer. There is a dropout performance in the network to prevent over- fitting [7,15].

ReLU was used to introduce non-linearity [14]. The stride is set to default. The input shape is 256x256x3 which means that colour image of size 256x256 should be provided to this network. ReLU performs activation function in this layer. First dropout layer added at the end of this layer which excludes 50 percent of the neurons to prevent overfitting. The output layer has 9 nodes corresponding to each classes of the hand gestures. This layer uses Sigmoid function [13] as activation function which results a likelihood value for each of the classes.

## 6. HIDDEN MARKOV MODEL

Hidden Markov Model are the most common models used for dealing with temporal data. It produces threshold model that yields the likelihood to be used as a threshold and helps in predicting the meaning of sequence of two

or more hand gestures.

## 7. EXPERIMENTAL RESULT

The model which was augmented with non-persistent data got 88.88 percent accuracy. It achieved this accuracy for the dataset in 100 epochs

The same dataset was moved as input to K Nearest Neighbors (KNN) models and multi class Support Vector Machine (SVM). These models establishes accuracy of 70.44 percent and 68 percent respectively. Major reasons for this is KNN and SVM have adaptability issue with non-linear dataset.

We have computed the Accuracy for each hand gesture using the standard accuracy formula

Accuracy = (TruePositive + TrueNegative)/Total (1)

And the obtained results are tabulated in Table 1 and the graph is plotted for the same, which is shown in Figure 3.

**Table I**

Accuracy Of Different Hand Gesture Classes

| Class No | Hand Gesture | Accuracy |
|---|---|---|
| 1 | Excuse Me | 86.7 |
| 2 | Good/Nice | 90.7 |
| 3 | Hi | 86.6 |
| 4 | I/Me | 85.6 |
| 5 | I Love You | 84.7 |
| 6 | Name | 94.7 |
| 7 | Night | 85.9 |
| 8 | What | 87.8 |
| 9 | You/Your | 96.8 |

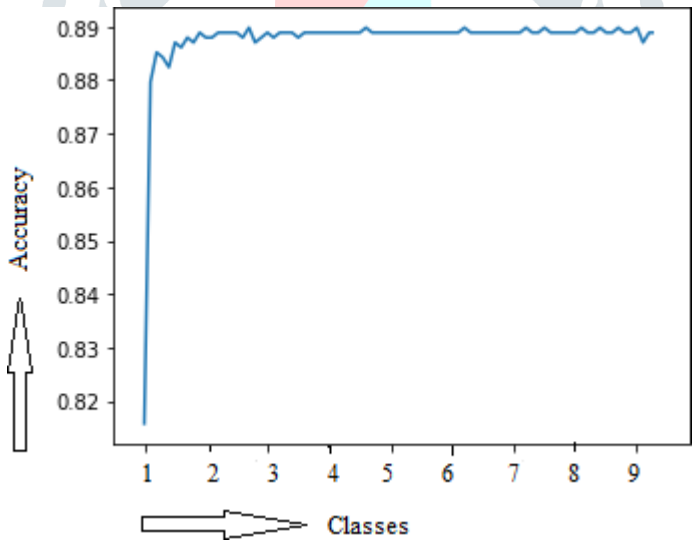



**Fig. 3.** Model Creation Phase.

## 8. CONCLUSIONS

In this work we have employed an hybrid effect of CNN into a HMM. Here, we explored the challenges in recognition of hand gesture. The capabilities of CNN and HMM has achieved an accuracy of 88.88 percent on our dataset. In this research work, we have considered only one hand gesture recognition but in future work, recognition of gestures made with both the hands can be considered and accuracy can also be improved using emerging techniques.

# REFERENCES


[1] R. Ul Islam, K. Andersson, and M. S. Hossain, "A web based belief rule based expert system to predict flood," in Proceedings of the 17th International conference on information integration and web-based applications services. ACM, 2015, p. 3.

[2] R. A. Dunne and N. A. Campbell, "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," in Proc. 8th Aust. Conf. on the Neural Networks, Melbourne, vol. 181. Citeseer, 1997, p. 185.

[3] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv preprint arXiv:1408.5093, 2014.

[4] Ravikiran Krishnan and Sudeep Sarkar. Conditional distance based matching for oneshot gesture recognition. Pattern Recognition, 48(4):1298–1310, 2015.

[5] Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Jan Kautz. Hand Gesture Recognition with 3D Convolutional Neural Networks. In Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1–7, 2015.

[6] Di Wu and Ling Shao. Deep dynamic neural networks for gesture seg- mentation and recognition. In Computer Vision-ECCV 2014 Workshops, pages 552–571. Springer, 2014.

[7] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper With Convolutions. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–9, Boston, Ma, USA, June 2015.

[8] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," Engineering Applications of Artificial Intelligence, vol. 22, no. 8, pp. 1141–1158, 2009.

[9] S. Miyamoto, T. Matsuo, N. Shimada, and Y. Shirai, "Real-time and precise 3-D hand posture estimation based on classification tree trained with variations of appearances," in Proceedings of the 21st International Conference on Pattern Recognition (ICPR '12), pp. 453–456, November 2012.

[10] A. Shimada, T. Yamashita, and R.-I. Taniguchi, "Hand gesture based TV control system—towards both user— machine-friendly gesture ap- plications," in Proceedings of the 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV '13), pp. 121–126, February 2013.

[11] C. Li and K. M. Kitani, "Pixel-level hand detection in egocentric videos," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13), pp. 3570–3577, 2013.

[12] Herve A. Bourlard and Nelson Morgan. Connectionist speech recogni- tion: a hybrid approach, volume 247. Springer Science Business Media, 1994.

[13] Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, and Hermann Ney. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In International Conference on Language Resources and Evaluation, pages 3785–3789, Istanbul, Turkey, May 2012.

[14] N. Pugeault and R. Bowden, "Spelling it out: real-time ASL finger- spelling recognition," in Proceedings of the IEEE International Con- ference on Computer Vision Workshops (ICCV '11), pp. 1114–1119, November 2011.

[15] V. Frati and D. Prattichizzo, "Using Kinect for hand tracking and rendering in wearable haptics," in Proceedings of the IEEE World Haptics Conference (WHC '11), pp. 317–321, June 2011.

[16] Vanishri Arun, Murali Krishna, Arunkumar B V, S K Padma, Shyam V., Exploratory Boosted Feature Selection and Neural Network Frame- work for Depression Classification, International Journal of Interactive Multimedia and Artificial Intelligence,2018.

[17] Vanishri Arun, Rubeena Banu, Shyam V., Meta-cognitive Neural Net- work Method for Classification of Diabetic Retinal images, IEEE Xplore 2016