Prediction of hand emoji using gesture recognition

Machine Learning (University of Delhi)

Prediction of hand emoji using gesture recognition

Aman Sundarka, Nandani Kothari, Mrunal Tare, Hrithik Basaiwal

Department of Computer Science, Smt. Kashibai Navale College of Engineering, Pune

***Abstract*--** **Emoji is a small digital picture or icon used to showcase an idea or a certain emotion. It exists in various genres, including facial expressions, common objects, places and types of weather and animals. This is a versatile representation of what a human body, emotions and environment contains of. Conversations through emoticon has become an adaptable way to have a conversation digitally. Interaction through emoji is a fun way of conversation starter and making the conversation a fruitful one. As it has become so popular and flexible, many a times, finding an emoji in the bunch of emoticons becomes highly tedious. To reduce this irritation a hand gesture prediction can be created where in users can create variety of hand gestures and the camera detects it and finds it for the user. This will not only make the process smoother but also more captivating. Emoji detector app would certainly show a mass of similar emojis detected by the hand gestures. The purpose is used to create a smarter communication, the emoji predictor changes the gestures based on the hand gesture variations. The classification is done using Convolution Neural network algorithm in order to avoid the feature extraction process. Deep learning is used for recognizing the images.**

I. INTRODUCTION

Emoji Prediction using is a fun variant to analyze sentiments. People use emojis almost every other day. They've become an addictive language which might more effectively express a plan or a feeling. This visual language is now an essential for online interaction, available not only on Twitter, but in other huge online platforms like Facebook, Instagram and all the other messenger apps like WhatsApp too. Currently, the keyboard on iOS can detect emojis but only based on certain keywords and certain images.

Majorly, even though its status as language form, emojis have been till today religiously studied from a neuro linguistic communication Processing (NLP) standpoint. The connection between text-based messages and emojis remained virtually not so explored.

In this application project will look forward to filling this bridge by investigating the relation between words and emojis, learning the problem of predicting which emojis spring to mind by text-based tweet message. We construct classifiers that studies to join emojis with sentences. The models we prospect here is Multinomial Naive Bayes and Bidirectional LSTM. Standard Bag of Word TF-IDF and pre-trained GloVe model are used as word majorly. We work on humongous dataset of sentences with emoticons labels collecting from Twitter messages.

II. MOTIVATION

This study has the background of social cause. It tries to use hand gestures to help in decreasing the house abuse that women or rather anyone faces by making it easier for them to convey through emojis and hence makes it a lot more truthful, we submit a study which will pivot on these unhappy incidents that will reduce its probability alongside discovering a new approach to machine learning algorithms. The idea has definitely caused this application to come in its glory and resulting the desired output quickly to the user.

III. RELATED WORK

Emoticons, also known as ideograms or smileys, can be used as similar expressions of things, materialistic monuments and emotions. Being embedded d in Unicode, they have no language issues and are diffused on the Internet immensely. The acceptance of emojis has allured researchers from different research companies such as NLP, multimedia and data mining. The different non-verbal functions of emojis have a major role in their broad adoption to the amount that they can have their special linguistic reason with written text. Initiate emojis from a sentence is majorly a text classification problem. paper realizes the segmentation of hand gestures by establishing the skin color model and AdaBoost classifier based on hear according to the popularity of skin color for hand gestures, as well as the denaturation of hand gestures with one frame of video being cut for analysis. The real-time hand gesture tracking is also realized by Cam Shift algorithm.

Emotion analysis refers to use the NLP techniques to learn the descriptive information from a sentence. Currently, emotion analysis is a topic of huge interest and development, as it can be broadly used for getting reports from social media comments, survey responses and making data-future decisions. Some applications, such as recommender systems use binary classification to showcase if a user communicate. Through like or dislike of some product or service of a sentence. Other forms of emotion analysis use distinctive categorical labels quite same to the emoji tags to forecast on those different emotions, such as happy, sad, excited, quirky etc. Since emojis have multi-contextual constitution and is easily used in all the other languages, it distributes as a great emotion label that can summaries the variations in a sentence.

IV. GAP ANALYSIS

Studies in machine learning paradigm focus on improving accurate prediction based on test data sets. Various research papers on the solution of driver drowsiness are focused on improving algorithms primarily. While there was also a need to analyse this problem psychologically and then come across the solution.

While we are using the technique of convex hull process for recognition of hand gesture is practically lengthy. In general, “convex hull” cannot recognize the hand gestures in less time in previous studies. The prime focus area was palm of the hand and finding its centre of the palm and computing the most extreme points also using palms centre constructing a circle using maximum Euclidean distance as a radius in every research paper we reviewed. Performing AND operation between the threshold hand image and the circular ROI. By using this the finger slices used to be counted which could further resulting in number of fingers shown. So, we also incorporated a background subtraction algorithm along with motion detection, and PyCharm software. This provided us with mixed results, but overall accuracy was improved slightly. By using the background subtraction algorithm, we were able to refine the output to a higher level. Such approach has also never been used in current solutions available publicly.

Thus, we were successful in deriving tangible results from hand gestures and convert them to software input. Using all palms centre and fingers count land-marking points in conjunction we were also able to detect the hand position which indicated whether it was distracted or not.

V. PROPOSED WORK

Through the use of Machine learning models and regression analysis, the hand gesture will be detected. Convulation neural network (CNN) is a primary source of picture acknowledgement and pictures order. Subsequently we are going to recognize the hand gesture with the help of video sequence all the recognition sequence is going to be done by live video sequence. For distraction free recognition process, we are going to take the hand region alone by removing all the distractive parts and unwanted portions in the sequence of video. the process is done is steps- 1) Background subtraction 2) Motion detection 3) training the model. Background subtraction is used by programmers globally for detecting the hand motion of live video camera sequence. Using the background subtraction algorithm an average area of palm of the hand will be demarcated which will be used as the idle position of the hand. It will act as a hand alignment indicator. Such contraption has invariably improved the efficiency of application output. Motion detection compares the change in image pixels and conclude that the motion has occurred.

VI. GENERAL FRAMEWORK

1. Background subtraction

The main aim of the algorithm is to differentiate between foreground and background parts of the image. By using running averages concept, we make our system to look over particular 1200 frames after computing the running averages over the current frame and the previous frame. We bring in our hand to the background to let system know the difference between the hand and the background after figuring out the background. We calculate absolute difference between the background model and the current frame to recognize the news added object to the foreground that is our hand.

2. Motion detection

After monitoring the hand images and removing the unwanted parts from the image to focus on the fingers, the motion detection gets ready to work. The basic function of the motion detector is to when the moving objected is detected the video analyzers starts to analyze the change in position of the image that was captured previously. To do this, the concept of running averages is used. Our system looks over a particular scene for 30 frames. During this period, we compute the running average over the current frame and the previous frames.

Further quickly summarizing and reviewing the activity is done by motion detection in a case to identify unusual events or security breaches. Though the framework uses the neural responses in a Gaussian mixture model (NeRM) for background scenes and motion detection is performed results in improvement in efficiency is astonishing, which is clearly due to improved algorithm.

The proposed NeRM method uses the advantage of sparse synaptic connectivity.

3. Training the Model

The next step is to train our model. Using CNN model, we can train our data of hand gestures also CNN is the first level to recognize the input it uses the pixels of an image to study the image features using small grids of input data. Convulation is the mathematical operation that takes two inputs. Image matrix and filter.

The method of gesture recognition is divided into two categories:

1. Finding geometric features of the hand gestures, the use of the gestures of the regional structure characteristics and edge characteristics as the recognition feature. The learning ability in this is not strong. Sample size is increased, results in the rate of recognition is not improved much.
2. Recognition method based on convolutional neural networks. The complex non-linear mapping is achieved by this method. It is widely used in industry for gesture recognition, but the single hidden layer is not enough for the algorithm to learn features and the model can easily distracted and leads to local optimal solution.
3. Recognition based on hidden Markov Model identification method. Large number of state probability density is calculated by the HMM model and need to estimate the number of parameters more which results into slower recognition speed.

Datasets:

The dataset will be created at run-time only. It’s going to be a collection of 1200 images (which will later be predicted as emoji’s) after getting the gesture number as input and accordingly it’s going retrieve from the dataset for prediction.

Table 1 The training set is 10% of the total database

Category of actual gesture	Recognition rate (%)							
	G1	G2	G3	G4	G5	G6	G7	G8
G1	78.43	5.23	3.92	0	0	0	4.58	7.84
G2	1.96	81.05	15.69	0	0	0	1.31	0
G3	0.65	0	91.5	1.31	3.92	0	1.31	1.31
G4	0	0	9.8	80.39	7.19	0	1.96	0.65
G5	0	0	1.31	0	98.04	0	0.65	0
G6	0	0	0	0.65	0	96.73	1.96	0.65
G7	0	0	0	0	0	0.65	98.69	0.65
G8	0	0	0	0	0	0	0.65	99.3

Table 2 The training set is 30% of the total database

Category of actual gesture	Recognition rate (%)							
	G1	G2	G3	G4	G5	G6	G7	G8
G1	100.0	0	0	0	0	0	0	0
G2	5.88	87.50	6.62	0	0	0	0	0
G3	4.41	1.47	90.44	3.68	0	0	0	0
G4	0	2.94	1.47	95.59	0	0	0	0
G5	0.74	0	13.24	4.41	80.88	0	0.74	0
G6	0	0	0	0	0	100.0	0	0
G7	1.47	0	0.74	0	0	0	97.79	0
G8	0	0	0	0	0	0	0.74	99.26

Table 3 The training set is 50% of the total database

Category of actual gesture	Recognition rate (%)							
	G1	G2	G3	G4	G5	G6	G7	G8
G1	100.0	0	0	0	0	0	0	0
G2	1.18	96.47	2.35	0	0	0	0	0
G3	0	3.53	95.29	1.18	0	0	0	0
G4	0	1.18	0	97.65	1.18	0	0	0
G5	0	0	0	0	100.0	0	0	0
G6	0	0	0	0	1.18	98.82	0	0
G7	0	0	0	0	0	0	100.0	0
G8	0	0	0	0	0	0	1.21	98.79

VII. Results And Discussions

As there are eight types of gestures and number is different, for example, G3 has 1208 samples, G8 has 1070 samples, we just take 1070 samples randomly from each set of samples for experimentation. By adjusting the number of training sets and conducting experiments to compare recognition rates as in Tables 1, 2 and 3. We can see from Table 1 that when the number of training is small, the rate of recognition of a gesture is low. The network needs to be trained more times for achieving

better results. The reason for that is, there are various types of gestures in the database due to the shooting angle of the problem and that can be seen as a class of hand objects and for the similar kind of gesture there is some amount of distortion.

So, it hinders the weight of network into the stability point, so that when each training round ends, the mean square error of network would still be high. So, the network needs to be trained many times to lower the mean square error. The experimental results give that when training set is 30% of the total, the change in mean square error stays towards stability, and shown in Tables 2 and 3. Integrated recognition rate of all three experiments: 90.52, 93.93, 98.52%. For comparison purpose, Sebastien Marcel Static Hand Posture data set, here this data set has six gestures of ten people, with a total of around 5,000 samples. The results of experiments are then compared with the BoFSURF+SVM, BoF-SIFT+SVM and Skin color model +CNN [42] for the same data.

Selecting one of the 1000 images as a training sample, the other remaining 4000 images are being used as test samples. Since the iterative equations of BoF-SURF + SVM and BoF-SIFT + SVM are the same, the subsequent iterations are represented by BoF-SURF + SVM. These images are used as input for the network after conversion of size, data range changes and mean operation. The train loss and test loss for the three networks would be gathered. Data from experiment for comparisons of the three networks are shown in Figs. 1, 2. Table 1 has data for test accuracy. Hence, we can say that recognition rate was fairly improved by using the method stated in this paper. Also, from Fig. 1 we can see that the train losses of these three networks all tending to 0. In Fig. 2 we see that the test loss of the single channel Skin color model +CNN is tending to 0.356, the test loss of the single channel BoF-SURF + SVM gets to 1.071 also the test loss of the Paper algorithm falls around 0.331.

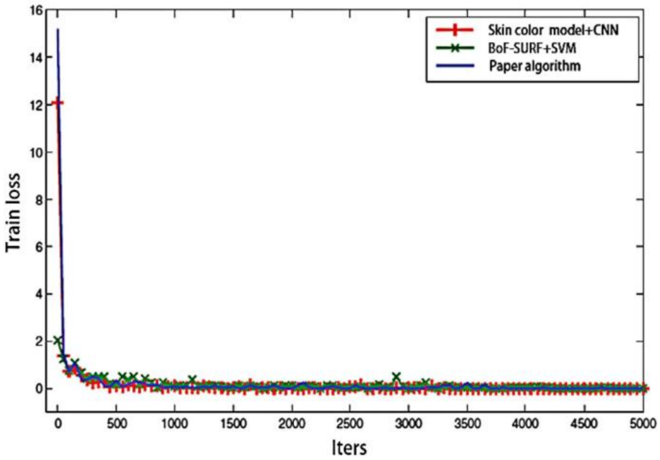

Fig. 1. The comparison of train losses

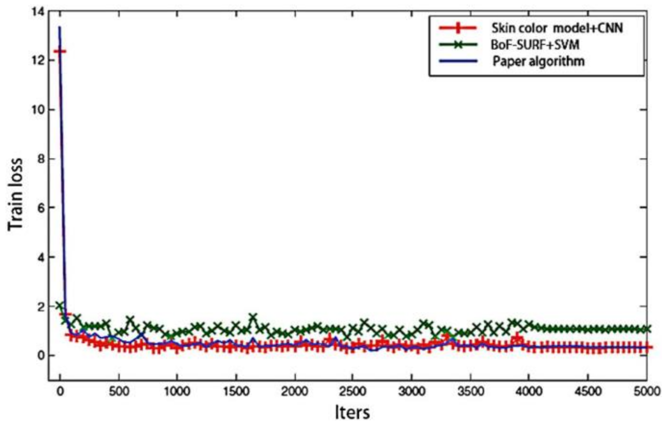

Fig. 2. The comparison of test losses

VIII. CONCLUSION

The following paper presents a hand gesture recognition method based on Convolutional neural networks for recognizing the gestures. We can extract the features, with the help of the layer of convolution. To avoid the sound judgement of selection of features artificially, considerable amounts of translation, scaling and rotation invariance are used. Also, the joint bilateral filter. By making use of the depth image for segmenting of colored image and removal of background noise, enhancing the robustness of the complete system. Along with accurate recognition of hand gestures experiments conducted have shown that this method is effective and favors in good results of convolutional neural networks for gesture recognition.

IX. REFERENCES

[1] Acharya, Shailesh, Ashok Kumar Pant, and Prashnna Kumar Gyawali. "Deep learning based large scale handwritten Devanagari character recognition." 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA). IEEE, 2015.

[2] Dessai, Brijeshwar, and Amit Patil. "A Deep Learning Approach for Optical Character Recognition of Handwritten Devanagari Script." 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). Vol. 1. IEEE, 2019.

[3] Xiao Niu Ching Y.Suen "A novel hybrid CNN–SVM classifier for recognizing handwritten digits"

[4] Kumar, Munish, et al. "Character and numeral recognition for non-Indic and Indic scripts: a survey." Artificial Intelligence Review 52.4 (2019): 2235-2261.

[5] Mukhopadhyay, Anirban, et al. "A study of different classifier combination approaches for handwritten Indic Script Recognition." Journal of Imaging 4.2 (2018): 39.

[6] Obaidullah, Sk Md, et al. "Automatic Indic script identification from handwritten documents: page, block, line and word-level approach." International Journal of Machine Learning and Cybernetics 10.1 (2019): 87-106..

[7] Deokate, Sarika T., and Nilesh J. Uke. "CNN Classification Approach For Analysis And Recognition Of Marathi Manuscript."

[8] Patil, Yamini, and Amol Bhilare. "Digits Recognition of Marathi Handwritten Script using LSTM Neural Network." 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2019.

[9] Deokate, S. T., and N. J. Uke. "Hybrid methods for Segmenting and Identifying the Marathi Text." 2019 IEEE 5th International Conference for Convergence in Technology (I2CT). IEEE, 2019.

[10] Gurav, Yash, et al. "Devanagari Handwritten Character Recognition using Convolutional Neural Networks." 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE). IEEE, 2020.