Hand Gesture Recognition Using Deep Learning

Soeb Hussain and Rupal Saxena
Department of Chemistry
Indian Institute of Technology, Guwahati
Guwahati, India
soeb.hussain@iitg.ernet.in and rupal.saxena@iitg.ernet.in

Xie Han, Jameel Ahmed Khan, Prof. Hyunchul Shin
Dept. of Electronics and Communication Engineering
Hanyang University
Sangnok-gu, Korea
xiehan@hanyang.ac.kr and shin@hanyang.ac.kr

***Abstract*—In order to offer new possibilities to interact with machine and to design more natural and more intuitive interactions with computing machines, our research aims at the automatic interpretation of gestures based on computer vision. In this paper, we propose a technique which commands computer using six static and eight dynamic hand gestures. The three main steps are: hand shape recognition, tracing of detected hand (if dynamic), and converting the data into the required command. Experiments show 93.09% accuracy.**

Keywords—computer vision, deep learning, hand gesture, neural network, transfer learning, hand gesture recognition.

I. Introduction

Gesture recognition is the mathematical interpretation of a human motion by a computing device. Modern research of the control of computers changes from standard peripheral devices to remotely commanding computers through speech, emotions and body gestures [1]. Our application belongs to the domain of hand gesture recognition which is generally divided into two categories i.e. contact-based and vision-based approaches. The second type is simpler and intuitive as it employs video image processing and pattern recognition.

The aim is to recognize six static and eight dynamic gestures while maintaining accuracy and speed of the system. The recognized gestures are to command the computer.

Division of hand gestures are explained in the block diagram shown in Fig. 1.

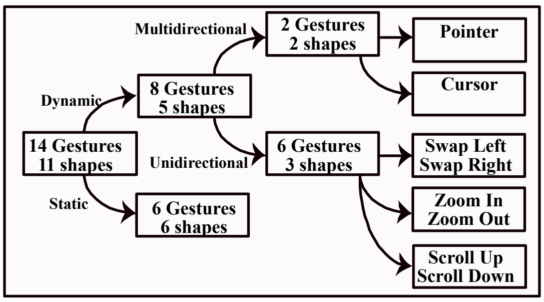


Fig. 1. Division of eleven shapes into fourteen gestures

For hand shape recognition, a CNN based classifier is trained through the process of transfer learning over a pretrained convolutional neural net which is initially trained on a large dataset. We are using VGG16 [2] as the pretrained model.

Each frame after resizing and padding is entered to the classifier. If the classified hand is a static gesture then it immediately passes to commanding phase. Otherwise, it passes to hand tracing phase. The block diagram of our proposed method is shown in Fig.2.

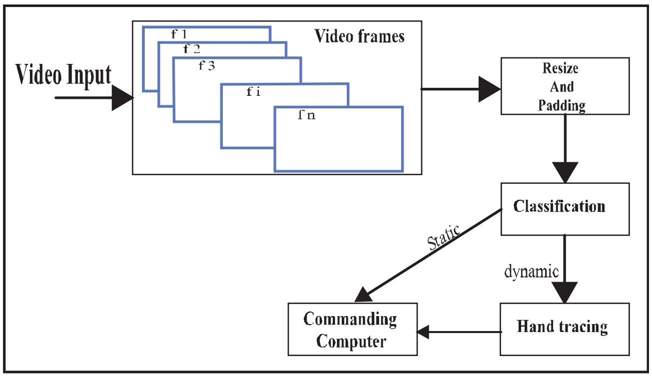


Fig. 2. Workflow

II. Hand Shape Recognition Using Transfer Learning

For hand shape recognition, the classifier is trained through the process of transfer learning [3] over a pretrained CNN that is initially trained on a large dataset.

Transfer learning is transferring learned features of a pretrained network to a new problem. The initial layers of the pretrained network can be fixed, the last few layers must be fine-tuned to learn the specific features of the new data set.

In our work, VGG16 a CNN architecture is used as the pretrained model. It consists of 13 convolution layers followed by 3 fully connected layers. A convolutional neural network (CNN) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex. We need to recognize eleven hand shapes, hence CNN is trained as a classifier using transfer learning method. To reach the desired output, network model needs to be altered. Therefore, two layers of the model were replaced with a set of layers that can classify 11 classes. All other layers remained unaltered. To avoid over fitting, the Regularization along with a more diverse dataset was introduced. Regularization involves modifying the performance function which is normally chosen to be the sum of the square of the network errors on the training set. The Classifier used over 55 thousand self-created image dataset out of which 70 percent were used for training and rest for testing. If recognized hand gesture is a dynamic hand gesture then it will further be traced to detect motion.

III. TRACING OF DETECTED HAND (IF DYNAMIC)

Recognition of a static gesture requires only the hand shape. Once hand shape is classified as static gesture by the trained classifier, command is given to the computer. Unlike static gesture, dynamic gesture requires both the hand shape as well as the motion of hand. For tracing dynamic hand gestures, hand area is segmented out using HSV (Hue, Saturation, Value) skin color algorithm in a frame, followed by cropping blob area. Centroid of the blob is detected and traced. The main idea in this stage consists in retrieving the coordinates of the traced hand's center in each frame. These coordinates will be used in order to know which computer command corresponds to which motion. Coordinates will be used differently for each gesture, depending on detected hand shape.

Five out of eleven hand shapes are used for dynamic hand gestures and rest for static hand gestures. These dynamic hand shapes are categorized into unidirectional and multi-directional hand gestures. Unidirectional hand gestures require shape and direction of motion of hand for commanding whereas multidirectional gestures require the position of hand along with its shape. Out of five dynamic hand shapes three are used for unidirectional gestures namely: swap, scroll and zoom, and remaining are used for multidirectional gestures of pointer and cursor. Each unidirectional gesture can further be used for differentiating two hand gestures depending on the direction of motion, e.g. swap can be left or right, depending on the direction of motion of hand.

Tracing involves extracting position of hand which is done by skin color detection, skin cropping, blob detection and centroid extraction. Hence tracing on the whole is a comparatively time consuming process. This process of tracing can be avoided after certain frames for unidirectional gestures as it only requires the direction of motion which can be derived from the few initial frames. Hence, the direction of unidirectional dynamic gestures can be determined by comparing centroid of initial frames.

IV. EXPERIMENTAL RESULTS

Our prototype was tested in different backgrounds by seven volunteers who did not train the system. Each of them performed all the hand gestures. We compared our results with CNN architecture AlexNet. Obtained results are shown in the Fig. 3. below. Overall accuracy for AlexNet is 76.96%. Our Recorded accuracy is 93.09%.

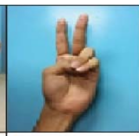
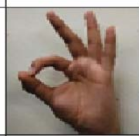
Gesture				
Command	Zoom	Double click	Left click	Cursor
Type	Dynamic	Static	Static	Dynamic
Alexnet	78.5%	42.8%	83.3%	91%
VGG16	100%	100%	80%	100%
Gesture				
Command	Chrome	Undo	Swap	Scroll
Type	Static	Static	Dynamic	Dynamic
Alexnet	100%	91.6%	57.1%	33.3%
VGG16	100%	100%	100%	66.66%
Gesture				
Command	Right Click	Pointer	File Manager	
Type	Static	Dynamic	Static	
Alexnet	100%	85.7%	83.3%	
VGG16	100%	100%	100%	

Fig. 3. Experimental results

V. CONCLUSION

We propose a vision based hand gesture recognition method using transfer learning. The method was made robust by avoiding skin color segmentation, blob detection, skin area cropping and centroid extraction for unidirectional dynamic gestures. Prototype was tested successfully on seven different volunteers at different backgrounds and light conditions with an accuracy of 93.09%.

ACKNOWLEGMENT

We would like to thank IC Design Education Centre (IDEC) at Hanyang University for Supporting tools and environment for this research.

REFERENCES

[1] Tahani Bouchrika, Mourad Zaied, Olfa Jemai and Chokri Ben Amar, "Ordering computers by hand gestures recognition based on wavelet networks" Communications, Computing and Control Applications, 2012. DOI 10.1109/CCCA.2012.6417911

[2] Karen Simonyan, Andrew Zisserman, "Very deep convolutional network for large scale image recognition" ICLR (Internaltional Conference on Learning Representations) 2015.

[3] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.