Effective Computer Vision Techniques for Real-Time Hand Gesture Recognition and Detection

Dr. Jayshree Pansare[1], Prof. Gajanan Aochar[2], Tejashree Salvi[3], Jessica Braganza[4], Akanksha Tiwari[5], Damini Kesharkar[6]

[1-6]*Modern Education Society's College of Engineering, Pune*

Abstract - *Gesture recognition technique is changing our relationship with technical devices by providing contactless or touchless user interface and opening the door to a whole new world of input possibilities. The main purpose of this paper is to explore the different gesture recognition and detection techniques. This paper describes research of gesture recognition and detection techniques that uses the knowledge of existing machine learning, deep learning, image processing and motion detection techniques. This also helps to advance the existing applications that a normal operating system provides by allowing users to interact with its applications with hand gestures which could be of their choice. After analyzing the research background for gesture control applications, we obtain that not all gesture control applications are cross-platform applications and lacking the high precision gesture recognition and detection techniques. The dataset would contain static and dynamic gestures. The important and reasonable 'Cross-platform' feature makes the operating system suitable to operate multiple portable devices such as routers, TV, mobile phones, Laptop/Desktop, Refrigerators etc. with hand gestures. Gesture based operating system could help people in different aspects and can be used in different fields such as healthcare, entertainment(gaming) and helps physically challenged people to enter into the world of technology.*

Key Words: Hand Gesture Recognition, Operating System, Machine Learning, Deep Learning, Image Processing, CNN, SVM, YOLOv3, OpenCV, cross-platform, static gestures, dynamic gestures.

I. INTRODUCTION:

Gesture is a way of expressing your emotions using actions. It involves body gesture, hand gesture, eye gesture, and many more. Hand and body gesture indicates a sign. The movement of the body or hand convey some message. This gives another way to how a human can interact with a machine.

In today's era, technology has taken a great control over humans. With the help of technology advancements, our lives have been simpler and faster. Research in Human Computer Interaction has given a tremendous boost to the advances in computer technology. Human-Computer Interaction plays a vital role in today's era of computer vision. HCI started off from keyboards, mouse and then to voice/speech recognition, gestures. Gestures being the most trending and popular way of interaction. This is because they are meaningful and communicative motions that helps to convey information and interacts with the environment in an efficient way. Gesture Recognition systems also decrease the requirement for handheld devices. Research on gesture-based technology started back in the early 1980s. The research started off to make interaction/control easier. For example, video games like the Sony PS3 imitates the user's action. Another example would be the vehicle gesture control, which allows the driver to control the vehicle.

Gestures can be either a part like facial gesture or hand/finger gesture or a whole-body gesture like standing, sitting, walking. Furthermore, gestures can be classified as dynamic or static. Static gesture includes the processing of a single image whereas dynamic gestures are a series of static gestures indicating a motion. They include processing of multiple images. It helps individuals to interact with the operating system by ease.

To distinguish different human gestures and utilize them for information transfer is a technique that needs to be developed.

In the gesture recognition method, a camera captures the user's body movements and transmits the data to a computer. These recognized gestures are used to control the operating system or variety of gadgets. Hand gesture recognition is a tricky task in Human-Computer Interaction fields. In order to be able to represent a serious alternative to conventional input devices like keyboards and mouse, applications based on computer vision are developed.

Hand gesture is not just an alternative to verbal communication, but it also benefits those lacking the ability to speak naturally. Many gesture recognition systems have been developed though not at an exceptionally large scale i.e., to its fullest capability, since this domain is still in research. Many techniques, tools, sensors, libraries, and algorithms are designed and developed for gesture recognition like glove-based sensing, convex-hull, skin detection algorithm and tools like OpenCV. The first gesture recognition hand tracking system, DataGlove, which is a glove-type device commercially available. Thus, using the above methodologies, a human-machine interaction interface can be developed that would enable individuals to remotely control a wide variety of devices through hand gestures.

Hand Gesture is popularly applied in fields such as: Medical where an application named Gesture makes

the doctors study the CT scans, X-rays and MRIs using gestures without scrubbing it. In automotives like BMW 7 Series, that has an inbuilt HRI system which allows the user to control the music system, the incoming calls and much more. Applications of hand gestures includes Robot controller, Sign language recognition, Augmented Reality (VR), Sign language detection, Computer Games, TV controller, mobile application and many more.

II. LITERATURE REVIEW:

A. Gesture Based Operating System Control:

The paper aims to propose a real time system using vision interaction and hand gestures using general-purpose hardware along with low-cost sensors. In this paper, RGB to HSV Conversion as well as Color Thresholding Technique is used. The frame is extracted, conversion of color model is done followed by color thresholding, detection of centroid and action according to selection. The result of the system is increased if limitations are minimized.

B. Hand Gesture Recognition Using Deep Learning:

The research in this paper aims to automate the interpretation of gestures using computer vision. Here, VGG16 which is a CNN architecture is used for pretraining the model which consists of convolution layers (13) followed by fully connected layers (3). To recognize the hand shapes, the CNN is trained with the help of transfer learning method. The Classifier used self-created datasets. Here, the training and testing dataset was divided into the ratio 70:30. If a dynamic hand gesture is recognized then the motion is traced and detect. To trace dynamic hand gestures; skin color detection followed by skin cropping along with blob detection and then centroid extraction is done. The blob's centroid is detected and then traced. For static gesture, the hand shape is classified by the classifier used for training. The result recorded 93.09% accuracy.

C. Power Point Control Using Hand Gesture Recognition Based on Hog Feature Extraction And K-NN Classification:

The aim of this paper is to enable users for managing and controlling the Slideshow Presentation using gestures. Here Skin Detection algorithm is used. After Resizing, Background Subtraction and Thresholding, a binary image is obtained with the help of HOG feature extraction. The image(gesture) is classified using K-NN Classification. The result acquired saw a drop in accuracy because of the lack of brightness and appearance of skin-colored objects other than the gestures.

D. A Smartphone-Based Application Using Machine Learning for Gesture Recognition:

The aim of this paper is to design a machine-learning model that is able to identify gestures through Hu image moments, because they are invariant to rotate, translate, scale, and have low computational cost. The images acquired are pre-processed. This is based on: 1) the segmentation of the image consisting of hand, in which the hand pixels are separated from the background 2) the binary conversion 3) enhancing the quality of the area segmented and the noise of the image is reduced and lastly, 4) the hand contour is retrieved after which it is possible to calculate Hu moments. Once the data is processed and normalized, they are compared using template-matching and a distance function to recognize the gesture. The conclusion observed was that the performance of the system is scalable by using powerful devices. However, good performances were obtained, showcasing a good identification capability.

E. Interactive Projector Screen with Hand Detection, Using Gestures.

In this paper, they proposed a system that is more portable and accurate than the old alternatives like speech recognition, laser-based gloves to control the operating system of the projector. They have implemented a virtual mouse which is a human-computer interface in a projector-camera detects the user's hand gestures (3 gestures) and replace mouse functions to control the projector by making an interactive wall that is variable in size (depends on the size of a projector). The System makes use of the OpenCV library for image processing and the C++ language for algorithm implementation and whole development. The System takes input through gloves by tracing its location. No external hardware and installation make this real-time system simple and efficient. The only constrain in this system is its region-specific feature and processing time as compared to other systems.

F. Gesture-based control for People Persons with Severe Motor Dysfunction by AAGI

They developed customized software, named the Augmentative and Alternative Gesture Interface (AAGI) which operates PCs and home appliances by utilizing a commercially available RGB-D camera by collecting data of different types of gestures (moving and depth images). Algorithms are made from scratch and camera libraries are used to obtain 2D and 3D images using shape information. 226 gestures were collected from 58 individuals with motor dysfunction and further voluntary movements were classified based on body parts. They developed 7 modules that can be handle by 7 different body parts (Finger, Head, Wink, Tongue, Shoulder, Knee, and Foot). The gesture setting function is used to assign gestures to operate electronic devices by

allowing users to set their own gestures to various operations. The given software provides 3 operations, 1. Remote controller 2. PC and 3. Calling. The Software uses Nature Remo which is a learning-type infrared remote controller to control home appliances. The main advantage of this software is its low-cost gesture-based interface. This system is in the testing phase and will soon release the English version (currently it is in Japanese language).

G. User-Dependent Gesture Recognition on Android Handheld Devices

They developed a user-dependent gesture recognition system, named GestWiz which controls mobile device using gestures. Data is recorded using accelerometer sensor readings. Accelerometer signals are translated into vectors of features in the process of feature extraction using a mixed feature scheme that combines "blind" statistical features with HoD (Histogram of Directions) feature descriptor. The model is trained using a distance metric learning algorithm and the metric within the system is trained by using an LMNN (Large Margin Nearest Neighbour). k-nearest neighbor (kNN) rule is applied for recognition function. The system was designed to run in real-time on a resource-constrained platform and therefore has low computational complexity, high computational, and storage efficacy. The system achieves an accuracy of 99.81%, which is higher than that achieved by using the Euclidean distance metric (93.39%).

H. Embedded Virtual Mouse System by Using Hand Gesture Recognition

This paper aims to implement an embedded virtual mouse system using hand gesture recognition. The Skin detection algorithm is used to detect a hand by determining the range of Y, U, V values to classify white and black pixels, the motion of the hand is determined using two continuous frames. Raster scan is used for labelling, Jarvis March algorithm finds the convex hull points. Detected gestures are replaced with the mouse function. The convex hull algorithm is performed well and the accuracy of the proposed system is 92%.

I. Android based American Sign Language Recognition System with Skin Segmentation and SVM

In this paper, an automation system is proposed that can convert sign language to text for easier interaction. The phone camera is used to capture the hand gesture that acts as an input to the system. Then the input image is processed and skin segmentation is carried out using the YCbCr System. Histogram of oriented gradients (HOG) is used for feature extraction. The output obtained after the HOG pre-processing is classified on the basis of training dataset using Support Vector Machines (SVMs) whose output is in the form of printed symbol on the screen. Then the final output is stored in cloud for future use, modifications without making changes in the system. The accuracy obtained by this system is 89.54%. But we may find variability in the types of acquisition devices, skin tone, background, illumination and minute gesture differences which makes the classification methods difficult to train and requirement of additional training for each captured image frame of a gesture increases.

J. Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language (2018)

In this model, Deep Convolutional Neural Network (DCNN) is used for extracting efficient hand features to recognize the American Sign Language (ASL) using hand gestures. The input gestures are collected using the webcam. Then the image is converted into a grayscale image. The human hand gesture images are taken for 26 alphabet signs of ASL for three different persons. The still hand image frame is captured from a running video frame and the convolutional neural network is used for feature extraction. Then the classification of the output data is performed by the MCSVM. The classification accuracy obtained is 94.57% which is significant for introducing the Sign Language Recognition of ASL for disable people as an output of HCI. For validation of this proposed model, our constructed dataset according to ASL conventions is used.

K. Learned Hand Gesture Classification through Synthetically Generated Training Samples

In this paper, A data generation technique is introduced that uses modern game engines to produce training data resembling the human hand that is realistic, rich but purely synthetic. Hand kinematics of a virtual avatar are manipulated to display hand gestures, while depth imagery is captured using Microsoft's AirSim plug-in. These captured depth samples undergo several levels of augmentation including scale, rotation and translation. These modifications are applied systematically and consistently across synthetic gestures. TensorFlow is used as the backend for creating neural network and training the model. Thus, the classifier was trained on synthetic samples for building robust training datasets providing insights for the performance improvement of learnt behavior derived from synthetic data.

L. YOLO9000 (2016):

This paper published a solution for the Image Processing challenge of scanning pixel by pixel, which requires tremendous computational power. YOLO looks through the image exactly once but in a smart approach. YOLO divides the image into a grid of 13 by 13 cells. Every cell is bound to predict five bounding boxes. YOLO produces some confidence score that describes how certain it is

the predicted bounding box surrounds any object. The confidence score on the bounding box plus the class prediction fuses into a final score that determines the probability that the bounding box contains the specific object type. The score represents the size of the box. YOLO model has trained on the PASCAL VOC dataset. There are 19 Convolutional layers, five max-pooling layers, including a SoftMax layer (DARKNET-19). It is one of the fastest algorithms for predicting objects within images.

M. LIGHT YOLO for high-speed gesture recognition:

This paper published an efficient model called Light YOLO. The model utilizes YOLOv2 for detecting objects. To enhance this gesture recognition ability on small-sized images, they introduced the Spatial Refinement Module moreover made extensive changes to YOLOV2 by removing the last max-pooling layer and sixth Convolutional layer for shaping network top features to higher resolutions. Light YOLO is a combination of YOLOV2 and Spatial Refinement Module. The research is beneficial to the researcher for performing experiments on the model by down-sampling some of the layers also exploring the Neural Networks. The accuracy of the model is 98.06%, which was measured using an F1 measure. Therefore, the accuracy curve is bending towards Light YOLO.

N. YOLOV3 (2018):

Some improvements were made in YOLOV2. Some of the major improvements includes YoloV3 has 53 CNN layers (Darknet53) stacked with 53 more layers producing 106 layers in total. Each CNN layer followed by Batch Normalization and Leaky RELU. There are no pooling layers, but convolutional layers are used instead. This prevents the loss of low-level features that pooling layers just exclude and thus gives the ability to detect small objects. The input image is down-sampled. While training, 3 bounding boxes are predicted, then anchors are used to calculate the Objectness score. To calculate anchors, K-means clustering algorithm is used. The probabilities of bounding boxes are extracted and an elementwise product of the Objectness score and list of confidence is taken. The maximum probability is selected among all probabilities; hence the object is found.

O. An Integrative Approach to Robust Hand Detection Using CPM-YOLOv3 and RGBD Camera in Real Time:

The researchers detect the position and orientation of the hand by YOLOv3 and Convolutional POS Machines. They created a HHD NET (Human Hand Detection Model) consists of 15 CONV layers, 6 max-pooling layers, 3 YOLOV3 layers. The model has trained on Oxford Hand Dataset. The images were captured by a digital camera in real-time, then human hands were predicted by a trained neural network. According to the results, it is clear that MS-RFCN [16] and other techniques have adequate accuracy, whereas the HHD Net model has a higher frame rate and average precision correspond to the other methods by using a low-end graphic card. It seems that the HHD Net is profoundly capable of accurately detecting the hand with orientation and position within an image.

III. CONCLUSIONS:

In this paper we compared the performance of Hand gesture recognition techniques and the analysis is based on important factors like number of objects, lighting condition, detection speed, frame size, etc. Following are the observations:

- CNN for static gestures gives an accuracy of 93.09% whereas DCNN with MCSVM gives an accuracy of 94.57% but with fixed unchanging background.
- HOG for feature extraction along with KNN and Euclidean distance metric for detection performs with an accuracy of 93.3%. However, HOG with SVM gives an accuracy of 89.54% but the performance is low considering the skin tone, background and lighting conditions.
- OpenCV is a tool but performance is region specific.
- On comparing YOLO, YOLO light and YOLO-V3, YOLO-V3 performs with higher accuracy and has the ability to detect complex objects.

REFERENCES

[1]. Prof. Yuvraj V. Parkale (2012). **Gesture Based Operating System Control**, 2012 Second International Conference on Advanced Computing Communication Technologies.

[2]. Soeb Hussain,Xie Han, Jameel Ahmed Khan, Prof. Hyunchul Shin and Rupal Saxena, **Hand Gesture Recognition Using Deep Learning**, IEEE 2017

[3]. Tejashree P. Salunke, S. D. Bharkad, **Power Point Control Using Hand Gesture Recognition Based on Hog Feature Extraction And K-NN Classification**, Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC)

[4]. Massimo Panella and Rosa Altilio**, A Smartphone-Based Application Using Ma-chine Learning for Gesture Recognition**, 2019, IEEE Consumer Electronics Magazine.

[5]. Sakshi Lahoti, Shaily Kayal, Sakshi Kumbhare, Ishani Suradkar, Vikul Pawar, **Android based American Sign Language Recognition System with Skin Segmentation**, 2018, ICCCNT.

[6]. Lindgren, Niveditha Kalavakonda, David E. Caballero, Kevin Huang, Blake Hannaford, **Learned Hand Gesture Classification through Synthetically Generated Training Samples**,2018 IEEE/RSJ

International Conference on Intelligent Robots and Systems (IROS)

[7]. Md Rashedul Islam, Ummey Kulsum Mitu, Rasel Ahmed Bhuiyan, Jungpil Shin, **Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language** ,2018 4th International Conference on Frontiers of Signal Processing.

[8]. Rishabh Sharma1, Raj Shikher, Nutan V. Bansode and Prachi R. Rajarapollu, **Interactive Projector Screen with Hand Detection Using Gestures,** 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)

[9]. Tsung-Han Tsai, Chih-Chi Huang and Kung-Long Zhang, **Embedded Virtual Mouse System by Using Hand Gesture Recognition,** 2015 International Conference on Consumer Electronics-Taiwan (ICCE-TW).

[10]. Tea Marasovic, Vladan Papi´c, **User-Dependent Gesture Recognition on An-droid Handheld Devices.**

[11]. Ikushi Yoda, Yuki Ozawa, Daisuke Nishida, Katsuhiro Mizuno, Kazuyuki Itoh, Tsuyoshi Nakayama, **Gesture-based control for People Persons with Severe Mo-tor Dysfunction by AAGI** ,2020 IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech 2020).

[12]. Redmon, J., Divvala S, Girshick R, Farhadi, A. (2016). **You Only Look Once: Unified, Real-Time Object Detection**. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[13]. Ni, Z., Chen, J., Sang, N., Gao, C., Liu, L. (2018). **Light YOLO for High-Speed Gesture Recognition**. 2018 25th IEEE International Conference on Image Processing (ICIP).

[14]. Joseph Redmon, Ali Farhadi (2018). **YOLOv3: An Incremental Improvement**.2018 Computer Vision and Pattern Recognition.

[15]. Muhammad Raza, Vachiraporn Ketsoi and Haopeng Chen (2019). **An Integrative Approach to Robust Hand Detection Using CPM-YOLOv3 and RGBD Camera in Real Time**. 2019 IEEE Intl Conf on Parallel Distributed Processing with Applications, Big Data Cloud Computing, Sustainable Computing Communications, Social Computing Networking.