Survey Paper on Vision based Hand Gesture Recognition

Simran Shah[1], Ami Kotia[2], Kausha Nisar[3], Aneri Udeshi[4], Prof. Pramila. M. Chawan[5]

[1,2,3,4]U.G. Students, Department of Computer Engineering, VJTI College, Mumbai, Maharashtra, India
[5]Associate Professor, Department of Computer Engineering, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *In the recent years, human computer interaction is becoming a vital part of most state-of-the-art technologies. The traditional mode of interaction via keyboards, mouse and joystick cannot meet the demands of this fast growing technology, and hence, in this paper, hand gesture recognition is studied and explained, to enable further development of natural communication between humans and computers. Different methods and algorithms that have previously been used in hand gesture recognition projects have been analysed and compared with respect to their advantages and drawbacks. Ongoing research challenges are also enlightened upon. Finally, a proposed methodology is discussed to show its increased efficiency in processing images, reduced training time and accurate hand gesture recognition.*

Key Words: Hand Gesture Recognition, Human Computer Interaction (HCI), Convoluted Neural Networks (CNN), Feature Extraction, Skin Color Detection, Classification

1. INTRODUCTION

Traditionally, users needed to tie up themselves with the help of wire in order to connect or interface with the computer system. In the previously used wired technology, a user cannot freely move as they are connected with the computer system with the wire and movement is limited with the length of wire. Instrumented gloves, which are also called electronics gloves or data gloves are an example of wired technology. These data gloves provide good results but they are extremely expensive to utilise in wide range of common application. Recently, some advanced vision based techniques have been introduced that require processing of image features like texture and colour.

The purpose of this project is to improvise natural interaction between humans and computers so that the recognised hand gestures can be used to convey meaningful information. We humans communicate not just with our words, but also with our gestures. With the recent development in computer vision and human computer interaction, we can create a system that is capable of identifying hand gestures and then performing suitable actions like moving the cursor on the desktop, opening certain applications, allowing smooth viewing and reading of PDFs, and also managing certain display settings.

We can define different positions or specified sequences of our hand movements as the hand gesture that our computer should recognise. Gestures may be static - requiring less computational complexity, or dynamic, which are more complex and also more feasible for real time systems. To exploit the use of gestures in HCI, it is important to provide the means by which they can be interpreted by the computers.

There are usually two main characteristics that should be deemed when designing an HCI system, and they are: Functionality and Usability. System functionality refers to the set of functions or services that the system equips the user to perform, and system usability refers to the level and scope under which the system can perform specific user purposes efficiently and more accurately.

1.1 Significance of Hand Gesture Recognition

With the quick development of hand gesture recognition systems, many different applications can be substantially improved. We may be able to control robots with simple hand postures, doctors and surgeons can use it to visualise more clearly each patient scan, various games can be simulated, online learning and teaching via distance learning techniques can be adopted and television sets can also be controlled. Amongst these, one of the most important would be sign language interpretation for the hearing impaired.

2. LITERATURE REVIEW

In [1], an efficient ISLR (Indian Sign Language Recognition) system is proposed which focuses on three main modules: segmentation, feature extraction and classification. Application of both Discrete Wavelet Transform (DFT) and nearest neighbour classifier is used to recognise the signs. It is observed that the nearest neighbour classifier provides 99.23% accuracy when used with the cosine distance metric. It has been proposed in [2] that the main step in hand gesture recognition is the segmentation of the hand from the whole image. Skin colour segmentation is applied to YCbCr space after conversion of all RGB colour space to YCbCr space. After Otsu Thresholding to separate background and foreground, they apply PCA to reduce dimensionality while preserving most information. A different method was proposed by the authors of [3]. Orientation Histogram is a method that allowed recognition of hand gestures even in less illuminated and complex backgrounds. The proposed method could not only segment the hand area but also extract the feature vectors from the gray scaled

motion images which represented 5 sign language words. A pair of data gloves are used as the sensing device to detect motion of hand and fingers presented in [4]. To recognise Korean Sign Language, an efficient classification of motions is proposed. For online pattern recognition, a fuzzy min-max neural network is adopted.

In [5], an overview of recent hand gesture systems is presented. The paper provides a review on camera interface. This paper explains briefly various concepts like modeling of gestures, localization and segmentation in image processing and colour detection. The paper mentions three learning algorithms: neural networks, Hidden Markov Models (HMM) and Instance-based learning for hand gesture recognition. The survey stated that the HMM technique for gesture recognition was found to be more accurate compared to neural network and instance-based learning.

In [6], a camera-based cursor control system is implemented which is a cost-free hand recognition software for laptops and PCs. By capturing the hand gestures from a webcam using a color detection technique, the system allows users to navigate cursor using their hand and perform left-click, right-click, dragging, cursor movement and file transferring. The methodology comprised of camera settings, capturing frames through web camera and masking techniques.

In [7], a media player is connected using a hand gesture interface. The authors propose four specific hand gestures: Play, Stop, Forward, and Reverse and four phases namely: Image segmentation, Feature Extraction and Classification. An artificial neural network utilized as a gesture classifier obtained an average classification rate of 95%.

In [8], authors have briefly described digital image processing techniques used for static hand gesture detection system. Scale Invariant Feature Transform (SIFT) classification method is used for hand gesture feature vector. The SIFT has been trained around edges such as scaling, rotation, addition of noise.

3. PROPOSED METHODOLOGY

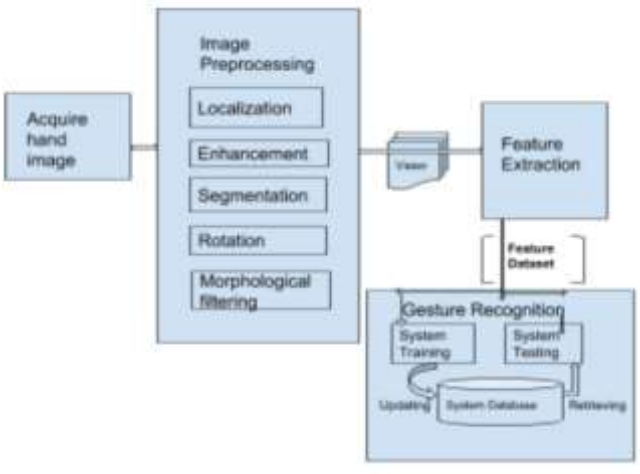

Fig -1: Architecture for proposed system

Step 1: Creating our Data Set:

We will be creating a data repository of gesture inputs with the help of a webcam. Several times in a minute, we will capture frames of the input and store and process it locally. This step eliminates the requirement of a hand glove, which is an expensive method, and also can prove to be a source of discomfort to the user.

Step 2: Defining the essential Hand Gestures:

From all the images that we have captured in our repository, we will now classify them as static, dynamic as well as unintentional movements.

Step 3: Image Processing:

The analysis of this input evolves in two sequential tasks. The first is the extraction of features from the raw image, and the second is computing the model parameters and essentially classifying the image. Further, skin colour detection plays an important role in the image processing.

The image processing step in our gesture recognition will involve the following steps: localization, image enhancement, segmentation, rotation and morphological filtering.

i. Localization - of the person performing the hand gesture from the rest of the image background. In order to allow smooth segmentation, it is necessary to determine the absolute position and orientation of the user.

ii. Enhancement - Image enhancement is done in order to improve illumination and remove blurring caused during image acquisition. Image features stand out more clearly with the use of this concept. We employ histogram equalisation for this purpose.

iii. Segmentation - Color based skin detection is most preferable for realistic applications. We use skin segmentation to reject as much of 'non-skin' background as possible. Since people with different complexion have different likelihood, an adaptive thresholding process is required to achieve the optimal threshold value. The output will be a grayscale image whose gray values represent the likelihood of the pixel belonging to skin.

iv. Rotation - PCA (Principal component Analysis) approach is utilized to rotate the segmented gesture image. This technique obtains a rotation invariant gesture image by coinciding the 1st principal component of the segmented hand gestures with the vertical axis.

v. Morphological filtering - Morphological filtering is necessary to be applied on segmented images to get a better smooth, closed and contour of a

gesture. This is achieved using a sequence of dilation and erosion operations over the rotation invariant segmented gesture image.

Step 4: Feature Extraction:

We now do data reduction using Feature extraction followed by feature analysis. The goal of feature extraction is to find the most discriminating information in the recorded images. The result of feature extraction is a description list or a feature vector.

A suitable selection of competent features is crucial to gesture recognition because hand gestures are not only rich in shape and variation but also motion and textures. Although hand postures can be recognized by extracting some geometric features such as fingertips, finger directions, and hand contours, these features are not always available and reliable because of self-occlusion and lighting conditions.

Explicitly specifying features is not easy. Therefore, whole images or transformed images are taken as input, and features are selected implicitly and automatically by the classifier.

Step 5: Gesture Recognition:

The resulting feature dataset will be then divided into training and testing dataset. Training set will be fed to a Machine Learning classifier. In the proposed system, a Convoluted Neural Network will be used. The model will train using training data set examples and modify hypothesis and weights to improve the accuracy of classification. Eventually, the best combination of weights, activation values and diverse data will create the most optimum model.

The hand gesture will now be given as input to the model to correctly classify it and accordingly action will be taken.

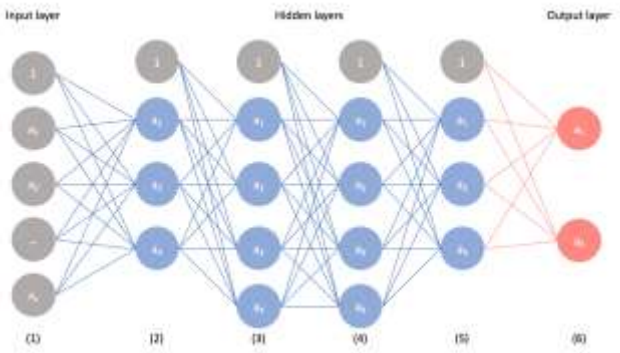

Fig -2: Generic Convoluted Neural Network

The metric we use for performance evaluation will be Accuracy. Where TN represents the true negative and FP represents the false positive and TP represents the true positive and FN represents the false negative. The accuracy is defined as the ability of the test to correctly identify a classified type with and without positives.

Accuracy (Ac) = (TP + TN) / (TP + TN + FP + FN) * 100

4. CONCLUSION AND FUTURE WORK

Hand gesture recognition is a method to allow seamless interaction between humans and computers, and that is the actual future of technology. We must ask ourselves what we can do to improve and maximize the accuracy and robustness of this system. More stress should be given to build up a dynamic gesture recognition system which gives satisfactory performance.

Without a doubt, there is always room for improvement in algorithms that are designed to learn by themselves. One of the major challenges we face in this system is the training time required for different models for recognising the hand gestures. Oftentimes, the memory requirements also increase manifold as the training data increases, and this causes a delayed result at run time. There is more scope in finding better hardware for data collection.

Our proposed methodology focuses on five main steps for an efficient hand gesture recognition. The system we are proposing aims to achieve superior results by focusing on a proper combination of preprocessing and classification tasks. We will evaluate our performance on the basis of Accuracy and Timeliness.

REFERENCES

[1] Mathavan Suresh Anand, Nagarajan Mohan Kumar, Angappan Kumaresan, "An Efficient Framework for Indian Sign Language Recognition using Wavelet Transform", Circuits and Systems, vol. 7, page 1874-1883, 2016.

[2] Mandar Kaue Ahuja, Amardeep Singh, "Hand Gesture Recognition using PCA", International Journal of Computer Science Engineering and Technology (IJCSET), vol. 5, issue 7, page 267-72, 2015.

[3] Hyung-ji Lee, Jae-Ho Chung, "Hand Gesture Recognition using Orientation Histogram", Proceedings of IEEE, Region 10 conference, 1999.

[4] Kim, J.S., Jang, W and Bien, Z. (1996), "A Dynamic Gesture Recognition System for Korean Sign Language", IEEE Transactions on systems, man and cybernetics, 26, page 354-359.

[5] Deepali N. Kakade, Prof. Dr. J. S. Chitode, "Dynamic Hand Gesture Recognition: A Literature Review ", International Journal of Engineering Research and

Technology (IJERT), ISSN: 2278-0181 Vol. 1 Issue 9, November- 2012

[6] Abhilash S, Lisho Thomas, Naveen Wilson, Chaithanya , "Virtual Mouse Using Hand Gesture", International Research Journal of Engineering and Technology (IRJET), Volume: 05, Issue: 04, e-ISSN: 2395-0056, p-ISSN: 2395-0072

[7] Hamid A. Jalab, Herman. K. Omer, "Human computer interface using hand gesture recognition based on neural network", IEEE 06 August 2015 , Electronic ISBN: 978-1-4799-7626-3, CD-ROM ISBN: 978-1-4799-7625-6

[8] Sagar P.More, Prof. Abdul Sattar, "Hand Gesture Recognition System For Dumb People", A R DIGITECH, International Journal Of Engineering, Education And Technology (ARDIJEET), ISSN 2320-883X, Volume 3 Issue 2, 2015