A Review: Hand Recognition System and Its Procedure

Tamanna¹ & Naveen Rawat²
M.Tech. Scholar¹, Asst. Prof. ECE of PPIMT²

Abstract—Hand gesture recognition possesses extensive applications in virtual reality, sign language recognition and computer games. The direct interface of hand gestures provides us a new way for communicating with the virtual environment. In this paper a novel and real-time approach for hand gesture recognition system is presented. In the suggested method, first, the hand gesture is extracted from the main image by the image segmentation and morphological operation and then is sent to feature extraction stage. In feature extraction stage the Cross-correlation coefficient is applied on the gesture to recognize it. Modification in algorithms has to be done to make it more efficient and accurate.

Keywords— ASL, HCI, RGB, HIS, MATLAB

I. INTRODUCTION

Recognition of sign language is one of the major concerns for dump and deaf people. Sign language recognition is a research area involving pattern recognition, computer vision, natural language processing. Sign language recognition is a comprehensive problem because of the complexity of the visual analysis of hand gesture and the highly structured nature of sign language. As well as it is considered as a very important function in many practical communication applications, such as sign language understanding, entertainment, and human computer interaction (HCI). Among natural human gestures occurring during non-verbal communication, pointing gesture can be easily recognized and included in more natural new human computer interfaces. The video streams of backgrounds are frequently influenced by the background changes such as illumination changes and changes due to adding or removing parts of the background. Therefore, the quality of the foreground and the segmented image of hand gesture severely drop.

Hand gesture recognition possesses extensive applications in sign language recognition, computer games and virtual reality. A gesture is a variety of non-verbal communication in which visible bodily actions may be used for communication. It can be classified into static and dynamic. The process of recognizing and predicting a gesture is known as Gesture Recognition and Sign language recognition is one of its applications. Sign language can involve combining orientation and movements of the hands, arms or body, hand shapes, and facial expressions to express thoughts and words that can be used for communication principally by Deaf and Dumb people. It can additionally offer an honest interface between computer and user, so in this paper we have a tendency to representing a hand gesture recognition system which will acknowledge most of the character from ASL with a smart accuracy. Figure 1 shows the American Sign Language symbols. In some passed decades Gesture recognition becomes very influencing term. There were many gesture recognition techniques developed for tracking and recognizing numerous hand gestures. Each one of them has their pros and cons. The older one is wired technology, in which users ought to traffic jam themselves with the help of wire so as to attach or interface with the computer system. In wired technology user cannot freely move in the room as they connected with the pc system via wire and limited with the length of wire. Instrumented gloves also referred to as physical science gloves or information gloves are the example of wired technology. These instrumented gloves made up of some sensors, provide the information connected to hand location, finger position orientation etc. through the use of sensors. These data gloves provide smart results however they are extraordinarily expensive to utilize in wide vary of common application.

Figure 1: American Sign Language Gestures are bodily actions made by humans to convey Meaningful information to others. It comes easily to human and so using these gestures as a mode of interaction will help humans to interact with computer easily. A gesture is scientifically categorized into two distinctive categories: dynamic
and static. A dynamic gesture is intended to change over a period of time whereas a static gesture is observed at the spurt of time. A waving hand means goodbye is an example of dynamic gesture and the stop sign is an example of static gesture. In this paper we mainly focus on hand gestures. The primary goal of hand gesture recognition system is to create a system which can identify specific hand gestures and use them to convey information or device control.

II. IMAGE SEGMENTATION

The goal of image segmentation is to cluster pixels into salient image regions, i.e., regions corresponding to individual surfaces, objects, or natural parts of objects. Segmentation could be used for object recognition, occlusion boundary estimation within motion or stereo systems, image compression, image editing, or image database look-up. Segmentation partitions an image into distinct regions containing each pixels with similar attributes. To be meaningful and useful for image analysis and interpretation, the regions should strongly relate to depicted objects or features of interest. Meaningful segmentation is the first step from low-level image processing transforming a greyscale or colour image into one or more other images to high-level image description in terms of features, objects, and scenes. The success of image analysis depends on reliability of segmentation, but an accurate partitioning of an image is generally a very challenging problem.

Segmentation techniques are either contextual or non-contextual. The latter take no account of spatial relationships between features in an image and group pixels together on the basis of some global attribute, e.g. grey level or colour. Contextual techniques additionally exploit these relationships, e.g. group pixels together with similar grey levels and close spatial locations.

It is a process in which we convert a RGB image or gray scale image into binary (Black and White) image. This is to be done because we can get only two objects i.e. black and white only in our image. Black with the background and white represents our hand. Otsu algorithm is used to convert image into binary. A good segmentation process is that process in which background doesn’t denote any part of hand and hand shouldn’t have any part of background. To obtain best result we have to choose best possible threshold value and segmentation can be done according to that value. The selection of the segmentation technique mainly depends on the type of image on which we have to do processing and Otsu algorithms had been tested and work efficiently with our hand gestures data. It is an unsupervised and nonparametric method of segmentation which can select threshold automatically and do segmentation.

- Non-contextual thresholding

Thresholding is the simplest non-contextual segmentation technique. With a single threshold, it transforms a greyscale or colour image into a binary image considered as a binary region map. The binary map contains two possibly disjoint regions, one of them containing pixels with input data values smaller than a threshold and another relating to the input values that are at or above the threshold. The former and latter regions are usually labelled with zero (0) and non-zero (1) labels, respectively. The segmentation depends on image property being threshold and on how the threshold is chosen.

- Adaptive thresholding

Since the threshold separates the background from the object, the adaptive separation may take account of empirical probability distributions of object (e.g. dark) and background (bright) pixels. Such a threshold has to equalize two kinds of expected errors: of assigning a background pixel to the object and of assigning an object pixel to the background. More complex adaptive thresholding techniques use a spatially varying threshold to compensate for local spatial context effects (such a spatially varying threshold can be thought as a background normalization).

- Colour thresholding

Color segmentation may be more accurate because of more information at the pixel level comparing to greyscale images. The standard Red-Green-Blue (RGB) colour representation has strongly interrelated colour components, and a number of other colour systems (e.g. HSI Hue-Saturation-Intensity) have been designed in order to exclude redundancy, determine actual object / background colours irrespectively of illumination, and obtain more stable segmentation.

- Morphological Filtering

The segmented images we get after applying the Otsu algorithm are not perfectly processed and further processing is needed in those images to remove unwanted data and errors. There are still some background parts which contain 1s and some hand parts which denote 0s. In order to remove that noise we have to apply morphological filtering techniques on those segmented images. It is necessary to remove these errors as they can create problem in recognition of hand gestures and reduce the system efficiency. So, morphological filtering is necessary to applied on segmented images and then we get a better smooth, closed and contour of a gesture. Dilation, Erosion,
Opening, and Closing is the basic operators that work in morphological filtering. Sample of preprocessing result is shown in Figure 2 and the experiments are performed in MATLAB. After the preprocessing we get a smooth and better hand gesture which can results a better efficiency.

![Gesture Images](image)

**Figure 2:** Result of the gestures after image preprocessing (a): main image (b): image after image segmentation (c): after morphological operation

- Cross-correlation Coefficient

Now we have to extract feature for gesture recognition. For feature extraction and matching we used Cross correlation Coefficient. In signal processing, cross correlation is a measure of similarity of two waveforms as a function of a time-lag applied to one of them. In this part we used this function for matching of hand gesture. The cross correlation coefficient is defined as Equation.

- Database Description

Our approach for hand gesture recognition is based on static mode so, our first problem is to gather a good quality of data since our classifier will classify characters according to it only. We had created our own database for each character of ASL which can includes 50 images i.e. 10 images for each gestures. During creating a database images captured should have uniform white color background that can be a white color rubber glove on hand as in contrast. We had done this in order to minimize noise and unwanted data so that we can easily do segmentation process. The user has to wear a black color cloth around his arm till wrist from the shoulder so that black color cloth can easily match with the background. Covered arm and the background should be of similar color. For static gestures only, the below shows the sample of our database that are more like a databases.

**III. LITERATURE REVIEW**

Hand gestures are powerful means of communication among humans and sign language is the most natural and expressive way of communication for dumb and deaf people. In this work, real-time hand gesture system is proposed. Experimental setup of the system uses fixed position low-cost web camera with 10 mega pixel resolution mounted on the top of monitor of computer which captures snapshot using Red Green Blue [RGB] color space from fixed distance. This work is divided into four stages such as image preprocessing, region extraction, feature extraction, feature matching. First stage converts captured RGB image into binary image using gray threshold method with noise removed using median filter [medfilt2] and Gaussian filter, followed by morphological operations. Second stage extracts hand region using blob and crop is applied for getting region of interest and then “Sobel” edge detection is applied on extracted region. Third stage produces feature vector as centroid and area of edge, which will be compared with feature vectors of a training dataset of gestures using Euclidian distance in the fourth stage. Least Euclidian distance gives recognition of perfect matching gesture for display of ASL alphabet, meaningful words using file handling. This paper includes experiments for 26 static hand gestures related to A-Z alphabets. Training dataset consists of 100 samples of each ASL symbol in different lightning conditions, different sizes and shapes of hand. This gesture recognition system can reliably recognize single-hand gestures in real time and can achieve a 90.19% recognition rate in complex background with a “minimum-possible constraints” approach [1].

Hand gesture recognition is a very challenging topic for real life applications because of its requirements on the robustness, accuracy and efficiency. This paper describes a system that enable a user to perform computer operations using hand gesture with a simple web camera as input device. This system involves four phases namely image acquisition, image pre-processing, feature extraction and gesture recognition. In the first phase, the input image is acquired with the help of a camera. In the second phase, the skin color of hand region is detected using HSV color space and morphological operations such as erosion and dilation are performed to remove noise followed by smoothing and thresholding of hand image. In Feature extraction phase, contours of hand image are detected. Finally, Gesture recognition phase includes recognizing hand gestures using contour analysis by comparing Auto-Correlation Function (ACF) amongst the contours and if they are close, then calculate Inter-Correlation Function (ICF) to truly determine similarity. Each recognized gesture is assigned with the corresponding action [2].

This paper describes a system that controls computer applications with the help of hand gestures. The proposed approach offers a unique method to identify hand gesture using contour analysis. Experimental
results show that the average recognition rate is 95%. The current system gives best result in a plain black background. Thus the current system puts lots of constraints on the user for successful working. The future work includes reducing these constraints so that the system is usable in more scenarios. Further enhancement of the technique proposed is possible using both hands for performing different computer operations. Experiments need to be done on a larger scale so that results can be more accurate [3].

Hand gesture is an active area of research in the vision community, mainly for the purpose of sign language recognition and Human Computer Interaction. In this paper, we propose a system to recognize alphabet characters (A-Z) and numbers (0-9) in real-time from stereo color image sequences using Hidden Markov Models (HMMs). Our system is based on three main stages; automatic segmentation and preprocessing of the hand regions, feature extraction and classification. In automatic segmentation and preprocessing stage, color and 3D depth map are used to detect hands where the hand trajectory will take place in further step using Mean-shift algorithm and Kalman filter. In the feature extraction stage, 3D combined features of location, orientation and velocity with respected to Cartesian systems are used. And then, k-means clustering is employed for HMMs code word. The final stage so-called classification, Baum-Welch algorithm is used to do a full train for HMMs parameters. The gesture of alphabets and numbers is recognized using Left-Right Banded model in conjunction with Viterbi algorithm. Experimental results demonstrate that, our system can successfully recognize hand gestures with 98.33% recognition rate [4].

Virtual environments have always been considered as a means for more visceral and efficient human computer interaction by a diversified range of applications. The spectrum of applications includes analysis of complex scientific data, medical training, military simulation, phobia therapy and virtual prototyping. Evolution of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient for the still widening spectrum of Human computer interaction. Gloves and sensor based trackers are unwieldy, constraining and uncomfortable to use. Due to the limitation of these devices the usable command set based diligences is also limited. Direct use of hands as an input device is an innovative method for providing natural Human Computer Interaction which has its inheritance from text based interfaces through 2D graphical-based interfaces, multimedia-supported interfaces, to full-fledged multi-participant Virtual Environment (VE) systems. Conceiving a future era of human-computer interaction with the implementations of 3D application where the user may be able to move and rotate objects simply by moving and rotating his hand - all without help of any input device[5].

IV. RESEARCH METHODOLOGY

Our approach for hand gesture recognition is consist of three step, I. Image segmentation; II. Morphological Filtering; III. Cross-correlation based feature extraction and matching. First of all we have to do preprocessing of data which is very important task in Hand Gesture Recognition system. Preprocessing should be done on images initially before we extract features from data of Hand Gestures. It is to be done to remove noise, unwanted errors and to make data efficient so, that it can be used for further image processing. We used two steps for preprocessing of data 1.Image Segmentation 2.Morphological Filtering.

V. CONCLUSION

Hand recognition is the term used for detection of Human based on the features. The Feature extraction and Feature Mapping is the main aspect to recognize the Gestures from the Database of features. Recognition of any individual is a task to identify people. Human recognition methods such as face, fingerprints, and iris generally require a cooperative subject, physical contact or close proximity.

REFERENCES

7. Cameron Stearns “A System for Cell Phone Anti-Theft Through Gait Recognition”, 2014

Imperial Journal of Interdisciplinary Research (IJIR) Page 1626
9. Muhammad Muaaz “Inuence of Walking Surfaces and Speeds on Accelerometer-Based Gait Recognition”, 2011