Efficient Hand and Finger Gesture Detection in Real-Time Using Skin Color Segmentation

Ravi Kumar[1] & Sandeep Singh[2]

[1]M. Tech Scholar, Computer Science, ABC University, New York, USA
[2]Associate Professor, Computer Science, ABC University, New York, USA

ABSTRACT

Gesture means an image of part of body. Gesture/posture Recognition is the technology that recognizes the posture of body. It recognizes hand, arms, head or any part of the body. So the goal of Gesture Recognition is to provide interface of human body with computer via mathematical algorithm. This paper gives a real time vision based method for recognizing human body image using MATLAB. A number of hardware techniques are used for gathering information about body positioning; typically either image-based (using cameras, moving lights etc.) or device-based (using instrumented gloves, position trackers etc.), although hybrids are beginning to come about. However, getting the data is only the first step. The second step, that of recognizing the sign or gesture once it has been captured is much more challenging, especially in a continuous stream. In fact currently, this is the focus of the paper.

KEYWORDS: Gesture, Feature, Vision, Image Processing

I. INTRODUCTION

Hand gesture recognition system received great attention in the recent few years because of its manifoldness applications and the ability to interact with machine efficiently through human computer interaction. In this paper a survey of recent hand gesture recognition systems is presented. Key issues of hand gesture recognition system are presented with challenges of gesture system. Review methods of recent postures and gestures recognition system presented as well. Summary of research results of hand gesture methods, databases, and comparison between main gesture recognition phases are also given. Advantages and drawbacks of the discussed systems are explained finally.

This research work focuses on the problem of gesture recognition in real time that sign language used by the community of deaf people. Research problem identified is based on Digital Image Processing using Color Segmentation, Skin Detection, Image Segmentation, Image Filtering, and Template Matching techniques.

This system recognizes gestures of ASL including the alphabet and a subset of its words.
The gesture recognition method is divided into two major categories a) vision based method b) glove based method. In glove based systems data gloves are used to achieve the accurate positions of the hand sign though, using data gloves has become a better approach than vision based method as the user has the flexibility of moving the hand around freely.

There are many possible vision based methods are available. Above all, Byong K. Ko and H. S. Yang developed a finger mouse system that enables a signer to specify commands with the fingers as in [2]. Apart from that, there are other different methods available such as colored hand-glove based method, Neural Network and PCA as in [3] to [5] etc.

In early days, for operation of robots a person should be present beside it. So it is necessary to design a robot which can be operated without a person being physically present. This paper is about operating a robot which is not near us and can be operated from a station; this is possible and is executed through a technique called IMAGE PROCESSING. In order to ease a feasible solution for user friendly interface, we can give commands to a wireless robot using hand gestures. Hand Gesture is frequently used in day to day life. The gesture recognition system consists of three stages image acquisition, feature extraction and decision algorithm. Once a gesture is recognized a command signal is generated and sent to micro controller. These command signals are then sent to the robot to operate it in a particular direction. One of the applications of the robot is to pick and place things where human intervention is not possible. The pick and place arm is a microcontroller based mechatronics system that picks the object from a source location and places it at a desired location, this action is controlled using a keypad switch. Further, the robot captures the video signals with its RF camera at 2.4GHz (mounted on the robot). The live video of the robot is received at the base station (computer). It is processed in the MAT Lab software and the video is processed in DSP code simultaneously.

The process of hand gesture recognition composes mainly of four stages:

- hand gesture images collection
- gesture image preprocessing using some techniques including edge-detection, filtering and normalization
- capture the main characteristics of the gesture images using feature extraction algorithms, and
- the evaluation (or classification) stage where the image is classified to its corresponding gesture class.

Real time gesture recognition is a challenging task. Gestures are extracted from video representing the signs performing static or dynamic gesture. Use of video bases information is increasing more and more. Due to this many research is done in the area of video processing. Key frame is very useful in the area of video abstraction, video summarization, video editing and animation. With the large amounts of video data
available, it has become increasingly important to have the ability to quickly search through and browse through these videos. For that Key frame is very useful for the users.

Application

- Hand gesture controlled robot for physically challenged.
- Hand gesture controlled doors and vehicles.
- Hand gesture controlled keyboard and mouse to interact with computer.
- Gesture controlled appliances like air conditioner.

Objective

The objective of this project is to find out the Key frame from the input video using skin color detection technique as complete moving picture in a video can be described as a finite image sequence, i.e., many still images. Each still image is expected as a single “frame”. All the frames in one video have a same size and the time between each two frames is equal, typically 1/25 or 1/30 seconds.

Color Segmentation

Color in an image is apparent by human eyes as a combination of R(red), G(green) and B(blue), these three colors i.e Red, Green and Blue are known as three primary colors. Other kinds of color components can be derived from R,G,B color represented by either linear or nonlinear transformations.
The RGB color components represent the incoming light, that is the brightness values of the image that can be obtained through (Red, Green and Blue filters) i.e RGB filters based on the following equations:

II. PROBLEM DEFINITON

Standard input devices like mouse and keyboards are used to provide inputs to the computer which may limit the systems capability and is not considered a natural way of communication.

As simple as it sounds the process of detecting a static hand gesture using a computer vision and from a set of predefined database might be very difficult. There has to be a clear idea of how to proceed towards the task of recognizing a static gesture. Therefore, some effort must be put to identify what might be the problems that may take place during the recognition process. Some of the methods that have been developed earlier helped in defining the problems in gesture recognition. As a real time video is used for static hand gesture recognition, the first and foremost problem is the one that arises with the question: How to extract gestures from a video frame? The success rate of recognizing a static gesture highly depends upon the environment in the video: the background of the video, presence of the objects other than the hand, and the lighting conditions; they play a vital role in the process of gesture recognition. However, Computer Vision and Image Processing provides techniques suitable for solving the problem of extracting gestures from a video. Identifying the hand which is the object of interest is also the problem. Thus, in online video when frame is triggered manually then in the identification process many unwanted noise may occur which tends to change the original form of the object of interest and thus may result in loss of information. Therefore, removing such unwanted noises becomes very important. Another problem that arises is the basic static hand gesture recognition identification which undergoes all the basic image pre-processing steps. Segmentation of the hand from the set of predefined static gestures and then identifying the gesture (set of numbers ranging from 0-9) is the most important step. Apart from this, tracking the fingers within the bounding box and its representation are some of the tasks essential for static gesture recognition. Classifying the gesture obtained from the video frame is the process that determines the outcome of the entire recognition process. This emphasizes on determining the features extracted (numbers from 0-9).

III. PROBLEM ANALYSIS

By analyzing the problem we observe that the main task is to find the feature points and the frames.

The main issue is to select proper image format for recognition of features from a set of recorded frames and proper video format for triggering a frame for identification purpose such that processing is efficient and fast enough.

- To identify the hand gesture irrespective of background and proper lighting condition is the main issue to be dealt with.
- The next task is to track the gesture for feature identification purpose from the manually triggered frame.
- The videos are taken from a low quality capture device, which consists of heavy noise due to low price image sensor. The next issue is to reduce the noise from the image such that minimum data loss can be achieved.
- When we have segmented the image consisting of human hands only, the next issue is to find the useful feature points from the image.
- Finally, the last task is to pass a set of frame database from all the basic stages of image processing.

IV. PROPOSED SOLUTION STRATEGY

Human hand gestures can be used to increase the human computer interaction in a new dimension and to decrease the cost of input hardware.

From a video containing static hand gesture frames are extracted and from those frames certain key frames are extracted based on a certain skin pixel range. Further, these frames as well as images which are captured through webcam are processed and goes through different phases like Image Processing, hand tracking, hand Gesture Recognition. The following sections will discuss the stages in detail.

- **Video processing**

For video processing, a video file will be taken which can be of any extension .mpg, .mpeg and .avi. Among the different set of videos, video file of choice is selected. Thus, the video is read using video reader which will read all the individual frames of the video and save each frame in a folder. In this project work a sample of video will normally generate 30 frames/second.

- **Image acquisition**

Image Acquisition is used to connect cameras to MATLAB which enables vision based hand gesture recognition. Image acquisition modes such as processing, hardware triggering, background acquisition, and synchronizing acquisition across multiple devices. After the image has been obtained, various methods of processing is applied to the image to recognize the hand gesture.

- **Key frame extraction**

From the extracted frames from a video the key frames are selected based on the skin color detection. Skin color is one of the most important features in the humans. There are lots of color spaces that have been used in early work of skin color, such as RGB, YCbCr, and HSV. Although RGB color space is one of the most used color spaces for processing images, it is not widely used in skin detection algorithms because the chrominance and luminance components are mixed. When building a system, that uses skin color as a feature for detection, several points must be kept in mind like what color space to choose and how to model color distribution.

In this project a skin color model based on HSV color space will be built because it has only two components (H, S) which help to speed up the calculations and also the transformation from RGB color space into HSV color space is done using simple and fast transformations. Thus the key frames will be selected if (Cb>=77 & Cb<=127 &Cr>=133 & Cr<=173). If the above condition satisfies the frame will be selected as key frame and further processed to give the gesture representing number.

- **Noise Removal**

To remove noise from the video the median filter technique is used. The median filter is a nonlinear digital filtering technique, often used to remove noise from an image or signal. Such noise reduction is a typical pre-processing step to improve the results of later processing (for example, edge detection on an image). The median is calculated by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value. Median filtering is very widely used in digital image processing because, under certain conditions, it preserves edges while removing noise.

- **Background Subtraction**

Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. Background subtraction is mostly done if the image in question is a part of a video stream. Background subtraction provides important cues for numerous applications in computer vision, for example surveillance tracking or human poses estimation [4].In the hand gesture recognition, the fingers are smaller than the palm so it is deleted using simple

morphology imerode. The palm is reconstructed i.e. little bigger than the original using imdilate. Finally the reconstructed palm is subtracted from the original palm which will provide only the fingers.

- **Image Segmentation**

In computing, segmentation refers to the process of partitioning an image into multiple segments. The main aim of segmentation is to simplify and/or change the representation of an image into something more meaningful and easier to use and analyze.

Image segmentation is one of the step involved in the process of gesture recognition in hand gesture recognition. The image captured by the camera cannot be used to track hand or recognize gesture as the image consists of other background objects and exists generally in RGB color space which makes skin detection process complex due to involvement of different color pixels in the image. So, in order to make skin detection process simpler, the image needs to be converted to a simpler color space which is easier to analyze and which involves lesser color pixels. Thresholding is the simplest method of image segmentation. For Hand gesture recognition multiband thresholding is used. As colorimages can also be thresholded, aseparate threshold for each of the RGB components of the image is designated. This reflects the way the camera works and how the data is stored in the computer, but it does not correspond to the way that people recognize colour. Therefore, the HSL and HSV colour models are more often used. The HSV color model is used in the static hand gesture recognition system.

- **Blob Analysis**

For image processing, a blob is defined as a region of connected pixels. Blob analysis is the identification and study of these regions in an image. The algorithms discern pixels by their value and place them in one of two categories: the foreground (typically pixels with a non-zero value) or the background (pixels with a zero value).The performance of a blob analysis operation depends on a successful segmentation of the image that is, separating the good blobs from the background and each other as well as eliminating everything else in the image that is not of interest. The acquired image may contain noise or spurious blobs or holes that may be caused by noise or lighting.

- **Feature Extraction and Classification**

Thus features, i.e. the blobs are extracted using region props within the boundary box. As the information about the contiguous areas are achieved using regionprops, it is then sorted and classified as the number of the gestures. It is labeled using labeling algorithm. Connected-component labeling is used in computer vision to detect connected regions in binary digital images.

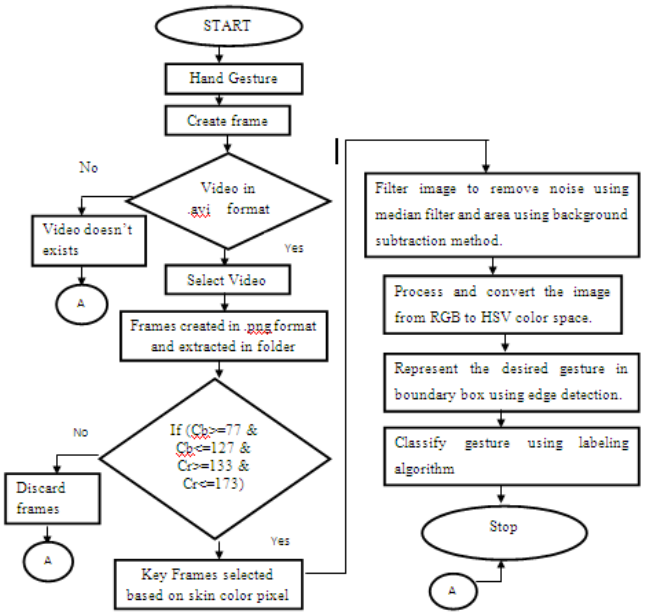

Fig 1: Flowchart for the proposed system

V. PROJECT ACHIEVEMENT

A static hand gesture recognition system was successfully implemented. In this project we explored the area of hand-gesture recognition for applications in human-computer interaction interface. After the completion of this project work it can be concluded that a few set of static gestures images can successfully undergo certain steps of preprocessing to display the number of fingers in the gesture. In real time static gesture recognition the two algorithms mainly Median Filter Algorithm and Contour Detection Algorithm was used for feature Extraction. Altogether the vision based method combines fast hand tracking, hand segmentation, region extraction and feature extraction to develop an accurate and robust hand gesture recognition method.

Future scope

This is an effective hand gesture recognition system to address the problem of extracting frames from a video and processing it. In the future scope, various hand gestures can be recognized and applied as input to the computer. The hand gestures representing numbers can also be converted into commands to perform related tasks in real time. Enhancing the recognition capability for various lightning conditions, which is encountered as a challenge in this project can be worked upon in future.

Conclusion

Variable numbers of key frames are extracted for different input video shot, which indicates the dynamic nature of the code. The code may be used in future for different purposes of processing videos. Key frames or the representative frames obtained are enough to represent the whole video. The main features of the video shot are grabbed efficiently in the extracted key frames which when summarized can represent the whole video in a concise manner. Relatively lower number of key frames are extracted which reduces the size of the summarized video. The main purpose of the proposed work is fulfilled since the efficiency of edge detection method is better than other methods. The summarized video is able to highlight the key contents of the original video shot. Image processing is also done to extract the fingers representing numbers

VI. REFERENCES

[1] **Xing-fang, Huang, & Jiang-she Zhang,** "Edge-Preserving Filtering for Grey and Color Image", In Proc. of the IEEE/Computer Science and Information Engineering

[2] **G. R. S. Murthy, R. S. Jadon,** "A Review of Vision Based Hand Gestures Recognition", International Journal of Information Technology and Knowledge Management, Volume-1, Issue-3, February 2012

[3] **Mandeep Kaur Ahuja, Dr. Amardeep Singh,"** A Survey of Hand Gesture Recognition", International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 5, May 2015

[4] **Yunyu Shi, Haisheng Yang, Ming Gong, Xiang Liu, and Yongxiang Xia,"**A Fast and Robust Key Frame Extraction Method for Video Copyright Protection", Journal of Electrical and Computer Engineering Volume 2017 (2017)

[5] **Nitin J. Janwe ,Kishor K. Bhoyar ,"**Video Key-Frame Extraction using Unsupervised Clustering and Mutual Comparison"

[6] **SwapnilAthavale ,Mona Deshmukh ,** "Dynamic Hand Gesture Recognition for Human Computer interaction; A Comparative Study" ,International Journal of Engineering Research and General Science Volume 2, Issue 2, Feb-Mar 2014

[7] **V. Pavlovis,R. Sharma and T. huang,** "Visual Interpretation of Hand Gesture for Human- Computer Interaction: A Review", *IEEE Transaction on pattern Analysis and Machine Intelligence, Vol. 19, No.7, Jul 1997.pp.,677-695.*

[8] **S. Zhao, W. Tan, C. Wu, L. Wen,** "A Novel Interactive Method of Virtual Reality System Based on Hand Gesture Recognition", *IEEE-978-1-4244-2723-9/09,pp., 5879-5882,2009.*

[9] **E. Stergiopoulou, N. Papamarkos,** "A New Technique for Hand Gesture Recognition",IEEE-ICIP,pp., 2657-2660,2016.

[10] **C. Burande,R. Tugnayat, N. Choudhary,** "Advanced Recognition Techniques for Human Computer Interaction", IEEE, Vol 2 pp., 480-483,20105