"GESTURE BASED WIRELESS CONTROL OF ROBOTIC HAND USING IMAGE PROCESSING"

Satyam M Achari[1], Shashwat G Mirji[2], Chetan P Desai[3], Mailari S Hulasogi[4], Sateesh P Awari[5]

[1,2,3,4] *Student, Dept of E&CE, GSIT, Karwar, Karnataka, India.*
[5] *Assistant Prof, Dept of E&CE, GSIT, Karwar, Karnataka, India.*

Abstract - *In this world many people are known to gestures, which is one of the most powerful communication technique amongst mediums. People say action speak louder than words. Communicating with gesture is powerful method. Many industrial places and home purpose robots are controlled by remote controllers. In our gesture control project, a particular robotic arm is controlled through hand gestures. Here, gesture provides a way of giving a speech by expressing an idea. Image processing is one the effective method for gesture recognition, which is use to process image signals. The purpose of this gesture recognition method is to capture a specific human hand gesture and performs applications depending on the user. Sign language recognition from hand motion is a useful aspect in gesture recognition research Human Computer Interaction (HCI). From the use of wireless communication, a robot can be control from a long distance by the user. This project aims in using hand gesture method and wireless communication to control a robot arm from a longer distance without using any physical remote controller.*

Key Words: **Image processing, wireless communication, HCI, hand gesture, image recognition.**

1. INTRODUCTION

A gesture is a communication that provides gestural messages by which body action takes place. Gestures are useful communication by humans. Hand gesture or hand motion gives a well-defined corresponding sensation to express one particular message. Information through hand gestures gives conversation in spatial and time-based manner. The term gesture means "the use of motions of limbs or body as a means of expression that emphasizes an idea". The main motive of hand gesture is to recognize a specific gesture and gives particular information depending on the type of gesture that has been provided. Representing and recognizing a gesture is a very difficult task because gestures occur in dynamically depending on shape and duration. Hand motion is active in gesture recognition research for Human Computer Interaction (HCI) [1]. Hand gesture provide a more natural-computer technology interface.

A robot is electro-mechanical system, basically a motion-based machine which is done by electronic circuitry and computer programming. The main goal of our project is to make Human Computer Interaction (HCI) system that has the capability to recognize certain hand gesture of humans to transform them into specific commands. Here, these commands are used by our robotic arm to simultaneously perform the required task. The entire task is done in real-time with fast and efficient algorithms. This prototype model is can be use in daily life with simple and efficient to use. Where in this kind of robotic project, the many applications mostly depend on the user i.e., the capacity or limit depends on how good the user use the system.

2. LITERATURE SURVEY

Gesture recognition enables humans to interface with the machine (HMI) and interact naturally without any mechanical or electro-mechanical devices. Using the idea of gesture recognition, we can point a finger at the computer screen so that the cursor will move accordingly. This could in theory make conventional input devices such as mouse, keyboards and even touch screens out of work. Gesture recognition can be conducted from computer vision and image processing. Most of the complete hand interactive systems can be considered to be comprised of three layers: detecting, tracking and recognizing. The detection layer is responsible for defining and extracting visual features that can be attributed to the presence of hands in the field of view of the camera. The tracking layer is responsible for accomplishing temporal data association between successive image frames, so that, at each instant in time, the system must be aware of "what is where". Moreover, in modal-based methods, tracking also provides a way to maintain estimates of typical parameters, variables and features that are not directly observable at a certain moment in time. Lastly, the recognizing layer is responsible for grouping the spatiotemporal data extracted in the previous layers and assigning the resulting groups with labels associated with particular classes of gestures.

We can use DSP processors other than the ordinary microcontroller, but to make the project cost-effective, it's not a favorable option. DSP processors are highly qualified for signal processing but on the other hand, the starter kits are high cost. Here, in our system we are using Raspberry PI, which is a low cost single board computer. We are writing the code in python where we are using OpenCV library instead of MATLAB software. Here, OpenCV is an open source library whereas MATLAB is costly. OpenCV uses less resources as compared to MATLAB. The Arduino UNO is an open source microcontroller which has embedded ATMega chip rather than using a simple ATMega microcontroller.

In today's world, some robots work with wired and some robots works wirelessly with RF modules. Here we are using controlling the robot wirelessly since it provides better portability. The RF module used here is nRF24L01 which is a low cost, ultra-low power transceiver that is used for wireless control and monitoring applications. Here, nRF23L01 is a device that can transmit data over long distances and has features like low power modes, multiple channels, channel hopping, frequency calibration, CRC, retransmit and so on. The nRF24L01 has a data rate of 1Mbps and 2Mbps speed.

3. WORKING PRINCIPLE

The whole working system is divided into two parts namely, a Controlling Unit (Transmitting Unit) and an Assembling Unit (Receiving Unit).

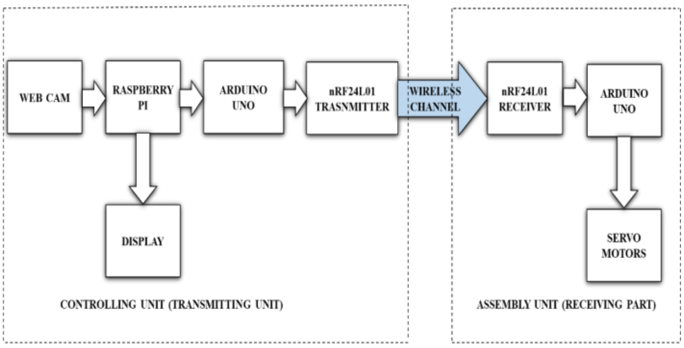

Fig -1: Block diagram of proposed system.

In this project, controlling of robotic arm is done via hand gesture through object recognition algorithm. The image of the hand is captured by web cam which is interfaced with the raspberry pi. The image of the hand is appeared on the display, so that the user may view and understand the image. The information from the Raspberry PI sent to the transmitting unit of the Arduino UNO and sends the information wirelessly to the receiving unit of the Arduino UNO. Here, wireless communication is done via nRF24L01 transceiver module, which is connected to both transmitting and receiving unit of the Arduino UNO boards. The information that is received by the Arduino UNO is processed which is used to control the servo motor which is attached to the robotic arm.

3.1 Controlling Unit (Transmitting Unit)

First the input image is taken via webcam on the Raspberry PI. This input image signal is then processed through OpenCV library where the program is written in python language. This programming code is assigned to a particular command to the input image signal. For example, we are assigning a particular hand gesture to the command. These assigned commands are exported from the Raspberry PI and uses Arduino UNO to send information through RF transceiver (in this case nRF24L01). This RF transceiver is present in both controlling unit and in assembling unit. The main function of Controlling Unit is to assign commands and to control the Robot Assembly Unit.

3.2 Assembly Unit (Receiving Unit)

The main function of this unit is to receive the command signals from the controlling unit to drive the robotic arm. The RF transceiver (nRF24L01) in this unit is to receive assigned commands from the controlling unit. The RF transceiver is connected to the Arduino UNO board where the control commands are received. The hand gesture commands are received by Arduino UNO board and as per the command the robotic arm will move. The microcontroller board (Arduino UNO) will decode the particular commands to control the servo motor. The code embedded in Arduino UNO board is written in Arduino language to control the servo motor.

4. GESTURE RECOGNITION

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Here, the gestures can originate from any motion of the body or commonly originate from the face or hand. Most of the gesture recognition are from face or hand gesture. Users can use the hand gesture technique to interact with the system or to control certain devices without physically touching them. Most of the hand gesture technique uses camera and computer vision algorithms to interpret sign language. Here, some identification and recognition of posture, gait and proxemics, and even human behavior are also subject of gesture recognition techniques. In our work we are using hand gesture recognition technique. Hand gesture recognition can be classified into two types:

1. Static Gestures
2. Dynamic Gestures

A static gesture is a specific hand configuration and pose, represented by a single image while a dynamic gesture is gesture motion, represented by a sequence of images. We will focus on the recognizing of a static gesture [2].

4.1 Gesture Recognition Process

There are two basic approaches in static gesture recognition, as described below.

[1] The top-down approach, where a previously created model of collected information's about hand configurations is rendered to some feature in the image coordinates. Comparing the likelihood of the rendered image with the real gesture image is then used to decide whether the gesture of the real image corresponds to the rendered one.

[2] The bottom-up approach, which extracts features from an input image and uses them to query images from a database, where the result is based on a similarity measurement of the database image features and the input features.

The disadvantage of the first approach is that it seems to use a high computational effort in order to achieve robust recognition. The second approach however requires an adequate pre-processing in order to achieve a reliable segmentation. This report mainly keeps the focus on the latter approach since this seems to be the commonly used one.

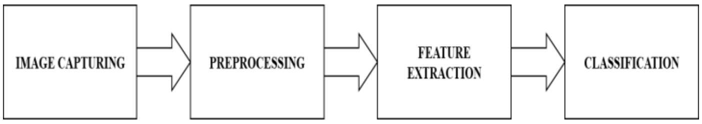

Fig-2: Schematic view of gesture recognition process

The whole process of static gesture recognition can be coarsely divided into four phases, as shown in Fig 1. Each phase performs a specific task, whose result is passed to the next phase. The commonly used techniques for each phase are described in the following subsections.

4.1.1 Image Capturing

The task of this phase is to acquire an image, or a sequence of images (video), which is then processed in the next phases. The capturing is done mostly by using a single camera with the frontal view of the person's hand, where gestures are performed. However, there also exist systems that uses two or more cameras in order to acquire more information's about the hand gesture.

In general, the following phases of the recognition process are less complex if the captured images do not have cluttered backgrounds, although several recognition processes seem to work in a cluttered image. Therefore, the images are captured in a cleaned-up environment having a uniform background. It is also desirable to have an equalized distribution of luminosity in order to gather images without shadowy regions.

4.1.2 Preprocessing

The basic aim of this phase is to optimally prepare the image obtained from the previous phase in order to extract the features in the next phase. The optimal result depends mainly on the next step, since some approaches only need an approximate bounding box of the hand, whereas others need a properly segmented hand region in order to get the hand silhouette. In general, some regions of interest, that will be subject of further analysis in the next phase, are searched in this phase. The most commonly used technique to determine the regions of interest is edge detection and skin colour detection.

4.1.3 Feature Extraction

The aim of this phase is to find and extract features that can be used to determine the meaning of a given gesture. Such a feature, or a set of such features, should uniquely describe the gesture in order to achieve a reliable recognition. Therefore, different gestures should result in different, good discriminable features. Furthermore, shift and rotation invariant features lead to a better recognition of hand gestures even if the hand gesture is captured in different angle.

4.1.4 Classification

The classification represents the task of assigning a feature vector or a set of features to some predefined classes in order to recognize the hand gesture. In previous years several classification methods have been proposed and successfully tested in different recognition systems. In general, a class is defined as a set of reference features that were obtained during the training phase of the system or by manual feature extraction, using a set of training images. Therefore, the classification mainly consists of finding the best matching reference features for the features extracted in the previous phase. This section presents an overview of the most commonly used methods in different hand gesture recognition systems.

5. IMAGE PROCESSING

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two-dimensional signals while applying already set signal processing methods to them. It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too [5].

Image processing basically includes the following three steps.

- Importing the image with optical scanner or by digital photography.
- Analyzing and manipulating the image which includes data compression and image Enhancement and spotting patterns that are not to human eyes like satellite photographs.
- Output is the last stage in which result can be altered image or report that is based on image analysis.

5.1 Purpose of Image Processing

The purpose of image processing is divided into 5 groups. They are:

1. Visualization - Observe the objects that are not visible.
2. Image sharpening and restoration - To create a better image.

3. Image retrieval - Seek for the image of interest.
4. Measurement of pattern – Measures various objects in an image.
5. Image Recognition – Distinguish the objects in an image.

5.2 Types of Image Processing

The two types of methods used for Image Processing are Analog and Digital Image Processing. Analog or visual techniques of image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The image processing is not just confined to area that has to be studied but on knowledge of analyst. Association is another important tool in image processing through visual techniques. So, analysts apply a combination of personal knowledge and collateral data to image processing.

Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

5.3 Fundamentals of Image Processing

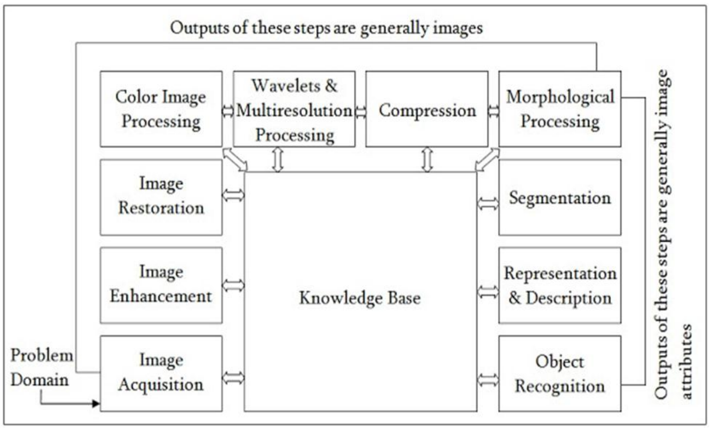

Fig-3: Fundamentals of image processing

5.3.1 Image Acquisition:

This is the first step or process of the fundamental steps of digital image processing. Image acquisition could be as simple as being given an image that is already in digital form. Generally, the image acquisition stage involves pre-processing, such as scaling etc.

5.3.2 Image Enhancement:

Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interest in an image. Such as changing brightness & contrast etc.

5.3.3 Image Restoration:

Image restoration is an area that also deals with improving the appearance of an image. However, unlike enhancement, which is subjective, image restoration is objective, in the sense that restoration techniques tend to be based on mathematical or probabilistic models of image degradation.

5.3.4 Color Image Processing:

Color image processing is an area that has been gaining its importance because of the significant increase in the use of digital images over the Internet. This may include color modeling and processing in a digital domain etc.

5.3.5 Wavelets and Multi-Resolution Processing:

Wavelets are the foundation for representing images in various degrees of resolution. Images subdivision successively into smaller regions for data compression and for pyramidal representation.

5.3.6 Compression:

Compression deals with techniques for reducing the storage required to save an image or the bandwidth to transmit it. Particularly in the uses of internet it is very much necessary to compress data.

5.3.7 Morphological Processing:

Morphological processing deals with tools for extracting image components that are useful in the representation and description of shape.

5.3.8 Segmentation:

Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually.

5.3.9 Representation and Description:

Representation and description almost always follow the output of a segmentation stage, which usually is raw pixel data, constituting either the boundary of a region or all the points in the region itself. Choosing a representation is only part of the solution for transforming raw data into a form suitable for subsequent computer processing. Description deals with extracting attributes that result in some quantitative information of interest or are basic for differentiating one class of objects from another.

5.3.10 Object recognition:

Recognition is the process that assigns a label, such as "vehicle" to an object based on its descriptors.

5.3.11 Knowledge Base:

Knowledge may be as simple as detailing regions of an image where the information of interest is known to be located, thus limiting the search that has to be conducted in seeking that information. The knowledge base also can be quite complex, such as an interrelated list of all major possible defects in a materials inspection problem or an image database containing high-resolution satellite images of a region in connection with change-detection applications [4].

6. EDGE DETECTION IN IMAGE PROCESSING

Edge detection includes a variety of mathematical methods that aim at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. The points at which image brightness changes sharply are typically organized into a set of curved line segments termed edges. The same problem of finding discontinuities in one-dimensional signals is known as step detection and the problem of finding signal discontinuities over time is known as change detection. Edge detection is a fundamental tool in image processing, machine vision and computer vision, particularly in the areas of feature detection and feature extraction.

The purpose of detecting sharp changes in image brightness is to capture important events and changes in properties of the world. It can be shown that under rather general assumptions for an image formation model, discontinuities in image brightness are likely to correspond to:

- discontinuities in depth,
- discontinuities in surface orientation,
- changes in material properties and
- Variations in scene illumination.

In the ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Thus, applying an edge detection algorithm to an image may significantly reduce the amount of data to be processed and may therefore filter out information that may be regarded as less relevant, while preserving the important structural properties of an image. If the edge detection step is successful, the subsequent task of interpreting the information contents in the original image may therefore be substantially simplified. However, it is not always possible to obtain such ideal edges from real life images of moderate complexity.

Edges extracted from non-trivial images are often hampered by fragmentation, meaning that the edge curves are not connected, missing edge segments as well as false edges not corresponding to interesting phenomena in the image – thus complicating the subsequent task of interpreting the image data.

Edge detection is one of the fundamental steps in image processing, image analysis, image pattern recognition, and computer vision techniques.

6.1 Edge Properties

The edges extracted from a two-dimensional image of a three-dimensional scene can be classified as either viewpoint dependent or viewpoint independent. A viewpoint independent edge typically reflects inherent properties of the three-dimensional objects, such as surface markings and surface shape. A viewpoint dependent edge may change as the viewpoint changes, and typically reflects the geometry of the scene, such as objects occluding one another.

A typical edge might for instance be the border between a block of red colour and a block of yellow. In contrast a line (as can be extracted by a ridge detector) can be a small number of pixels of a different colour on an otherwise unchanging background. For a line, there may therefore usually be one edge on each side of the line.

6.2 Contour Detection and Convex Hull

The simple Haar-like features (which are computed similarly to the coefficients in the Haar wavelet transform) are used in the Viola and Jones algorithm. The Haar-like features are robust to noise and various lighting condition because they compute the gray-level difference between the white and black area of rectangles. The noise and lighting variations affect the pixel values on the whole feature area. The integral image at the location of pixel [x, y] contains the sum of the pixel intensity values located directly above the pixel location [x, y] and at the left side of this pixel. So, A[x, y] is the original image and AI[x, y] is the integral image that is calculated by below equation 1:

$$AI[x, y] = \hat{I}A[x', y'] \text{.....................(1) } x' \leq x, y' \leq y$$

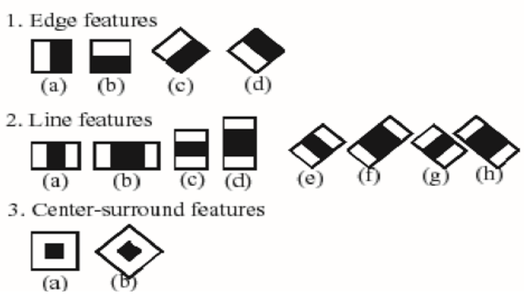

Fig-4: Common harr like features for integral images

Practically no single Haar-like feature can identify the object with high accuracy. However, it is not difficult to find out one Haar-like feature-based classifier that has better accuracy than the random guessing.

The AdaBoost based learning algorithm improves stage by stage overall accuracy, by using a linear combination of these individually weak classifiers. The AdaBoost learning algorithm initially assigns an equal weight to each training sample. We start with the selection of a Haar-like feature-based classifier for the first stage and got better than 50% classification accuracy. In next step this classifier is added to the linear combination with the strength that is proportional to the resulting accuracy. So, the training sample weights are updated i.e. training samples that are missed by the previous classifier are boosted in accordance. The next classification stage must achieve better accuracy for these misclassified training samples so that the error can be reduced. By this procedure we can improve the overall classification accuracy at further stage. The iteration goes on by adding new classifiers to the linear combination until the overall accuracy meets to the required level. At the final level the result is a strong classifier composed of a cascade of the selected week classifiers [3].

In Convex Hull algorithm first step is segmentation of the hand image that contains the hand to be located. In order to make this process it is possible to use shapes, but they can be changed greatly in interval that hand moves naturally. So, we select skin-color to get characteristic of hand. The skin-color is a distinctive cue of hands and it is invariant to scale and rotation. In the next step we use the estimated hand state to extract several hand features to define a deterministic process of finger recognition. After the hand is segmented from the background, a counter is extracted. The counter vector contains the series of coordinates of edges of hand. Then the processing of counter vector gives the location of the fingertip.

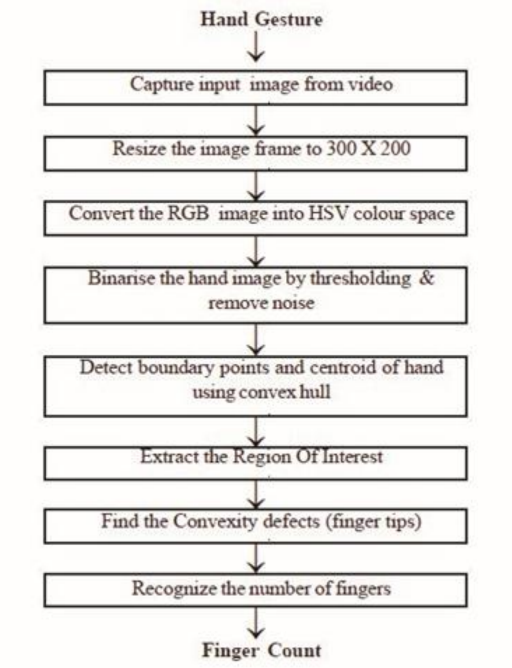

Fig-5: Flow of methodology for finger counter

In convex hull implementation firstly calculate the points with minimum and maximum x and y-coordinates and by joining these points a bounding rectangle is defined, within which the hull is contained. There will be other points of convexity too; we find the convex defects i.e. between each arm of hull. The defect points are most likely to be the center of the finger valleys. Then find out the average of all these defects which is definitely bound to be in the center of palm, but it is very rough estimate so average out and finds this rough palm center. Thus, the radius of palm is an indication of depth of the palm using radius. The ratio of palm radius to the length of the finger triangle should be more or less same, thus we can find out number of location of tip of finger. Here AdaBoost algorithm and Haar like feature set algorithm are adopted for hand detection and recognition. But when we combine every gesture in same program, it gets mixed and create confusion for real world hardware. So, we have changed the representation way of rectangle & do it with fingertip by connecting lines for clear visibility. This experiment we developed on open source library for computer vision application called Open Computer Vision Library (OpenCV).

7. RESULTS

The image in fig-6 shows the result of our project where a robotic hand is control through image processing wirelessly.

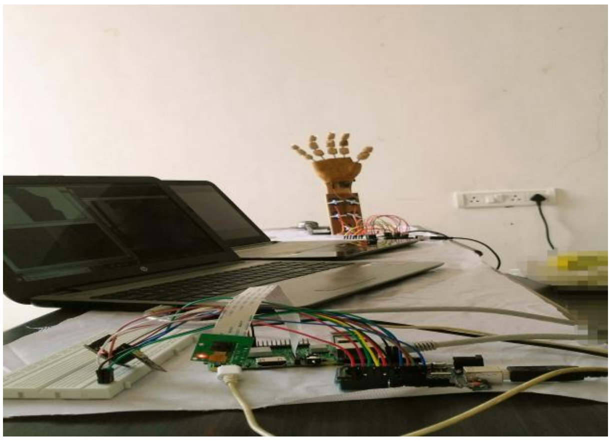

Fig-6: Image of the wireless control of robot hand through image processing

Fig-7 shows the binary image of the user's hand where the where it is used to find the contours.

Fig-7: Binary image of the hand.

Here, we have used Otsu's Binarization method. In this method, the OpenCV library automatically calculates the threshold value of the image from its image histogram. The resulted image is shown in fig-8.

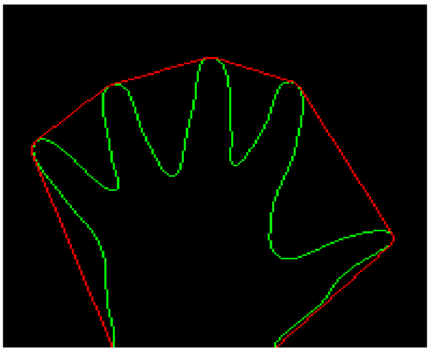

Fig-8: Image where contours are drawn.

8. CONCLUSION

By the use of images processing technique, a user is able to control a robot or any system. The use of gesture-based technique is useful in such kind of applications. Here, with the help of hand gestures-based approach dynamic application is possible, it means that the user will be able to provide multiple usage of the system without any limitations. These kinds of robotic applications are helpful many industrial and agricultural usage.

9. FUTURE SCOPE

1. Sign language recognition
   By the use of images processing sign language detection is possible where we can convert a symbol to text information.

2. Remote control
   From the use of hand gestures detection, the user can be able to control RC cars, drones etc.

3. From facial features
   Controlling a computer through facial gesture is possible via image processing where eye tracking is used in order to control the cursor.

4. In gaming technology
   Through the use of gesture-based control a player will be able to play games without using any controllers.

5. Alternative computer interfaces
   Instead of using mouse and keyboard by the use of hand gestures a user is able to interfaces with the computer by using only gesture-based control.

6. Full robot gestures
   By the use of full body gesture-based control the user will be able to control a humanoid robot.

REFERENCES

[1] Gaurav Chauhan and Prasad Chaudhari, "Gesture based Wireless Robotic Control using Image Processing", Dept. of Electronics and Telecommunication Engineering MIT Academy of Engineering, 26-28 Nov. 2015.

[2] Thierry Messer, "Static hand gesture recognition", Department of Informatics, University of Fribourg, 1700 Fribourg, Switzerland.

[3] Ruchi Manish Gurav and Premanand K. Kadbe, "Real time finger tracking and contour detection for gesture recognition using OpenCV", Dept. of Electronics and Telecommunication Engineering, VPCOE, Baramati, India.

[4] Anil K. Jain, "Fundamentals of digital image processing", Pearson Education, 2001.

[5] B. Chanda and D. Dutta Majumdar, "Digital image processing and analysis", PHI, 2003.