



Vision Based Hand Gesture Recognition for Robot Control

Pooja R. Maniar¹, Prof. R. C. Patel²

PG Student¹, Associate Professor²,

Instrumentation & Control Department^{1,2}

Gujarat Technological University^{1,2}

L. D. College of Engineering, Ahmedabad, Gujarat, India^{1,2}

Abstract: There is a drastic development in the human computer interfaces in order to facilitate users and make them comfortable. Our natural hand gestures can be used as an interface to operate machines, communicate with ingenious environments to control home appliances, robot, etc. In this paper, we proposed an effective algorithm for recognizes a set of static hand gestures for a robot control application. Our approach contains steps of pre-processing, skin colour detection and feature extraction. The main features used are of two types: appearance based features and the cross correlation based features. The result based on recognition rate was analysed.

Keywords: Hand Gestures, Pre-processing, Feature Extraction, Cross Correlation, MATLAB, Robot Control, skin colour detection.

I. INTRODUCTION

Hand gesture recognition is an important area of computer vision and pattern recognition field as the most flexible part of human body, hands play an important role in human's daily life communication [11]. The development of Gesture recognition is the witness of the human being effort aiming to communicate with electronic gadgets. From the wide range of applications, this paper focuses on the hand gesture recognition for robot control. What motivates us for this work is a robot navigation problem, in which we are interested in controlling a robot by hand pose signs given by a human [19].

Numerous techniques are used for the hand gesture recognition. Based on the extra devices required, we can classify it into data glove based technique, color glove based technique and the vision based technique. From all these three methods, vision based approach does not require any extra device to collect the raw information about the gesture. So we have used vision based approach in this paper.

Moreover, the recognition has to be done by one camera, so that one can operate as fast as he wants to. The sensor device used is an USB web cam. So, this makes it possible for any user to use it in his office or home. The system was developed using MATLAB R2014b on Windows 7 Operating System. The images after being captured by the web cam were saved in the database using .jpg format. The images are labelled using integer numbers starting from "1". The two databases were created using nine different images of single hand gestures. The process of hand gesture recognition would start with image processing techniques such as skin colour detection, noise removal, followed by feature extraction involves the classification of images based on the shape, area, position of centroid, etc.. Then additional classification is done by Euclidian distance, number of peak finder and template matching technique.

II. RELATED WORKS

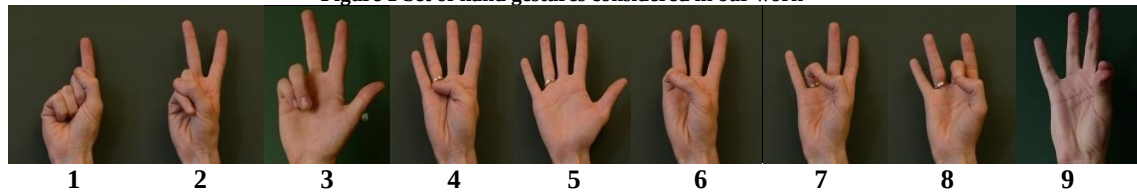
Early approaches to the hand gesture recognition problem in a robot control context involved the use of markers on the finger tips [1]. Hand gesture recognition is performed through a curvature space method in [2], which involves finding the boundary contours of the hand. This is an athletic approach that is scale, translation and rotation invariants on the hand appearance, yet it is computationally demanding. In [3], a vision-based hand pose recognition technique using skeleton images is implied, in which a multi-system camera is used to pick the centre of gravity of the hand and points with farthest distances from the centre, providing the locations of the finger tips, which are then used to obtain a skeleton image, and finally for gesture recognition. A method for gesture recognition for sign language interpretation has been proposed in [4]. In [5], the hand gestures are taken by a camera. Image transformations are carried out on the rgb image to convert into ycbcr image. The ycbcr image is transformed into binary image. This algorithm needs uniform and plane background. Edge detection algorithm is used to find the edges in the image. By making use of edge detection the orientation of hand is detected. The features like centroid, peaks detection, Euclidean distance and thumb detection are found. In [6], gives an algorithm for non-uniform background or 3D complex space, use of HMM based method to recognize the hand gestures. The input images are taken by a camera. Skin colour is used for segmentation. Other

computer vision tools used for 2D and 3D hand gesture recognition include specialized mappings architecture [7], principal component analysis [8], Fourier descriptors, neural networks, orientation histograms [9], and particle filters [10].

III. PROPOSED METHODOLOGY AND DISCUSSION

Consider a robot navigation problem, in which a robot responds to the hand appearance signs given through a human [18], visually observed by the robot by a camera. We are interested in an algorithm that enables the robot to identify a hand pose sign in the input image, as one of nine possible commands. The identified command will then be used as a control input for the robot to perform a certain action or execute a certain task [19]. For examples of the symbols to be used in our algorithm, see Figure 1. The symbols could be associated with various meanings depending on the function of the robot. For example, a “1.jpg” could mean “elbow up”, a “9.jpg” could mean “stop”. Furthermore, “2.jpg”, “3.jpg”, “4.jpg”, “5.jpg”, “6.jpg”, “7.jpg” & “8.jpg” could be interpreted as “elbow down”, “gripper open”, “gripper close”, “rotate clockwise”, “rotate anti-clockwise”, “shoulder up” and “shoulder down” respectively.

Figure 1 Set of hand gestures considered in our work



Our proposed method of hand gesture recognition consists of the following stages:

- Pre-processing includes tasks such as hand tracking from the entire image, removing the background noise and converting the original colour image into binary image.
- Skin colour detection based on learned skin colour statistics, producing a BW image output.
- Appearance based feature extraction include the features which involves the classification of images based on the shape, area, position of centroid, etc. Features included in the appearance based classification: 1. Eccentricity & Solidity ratio 2. Extent and 3. Compactness factor.
- Cross correlation based feature extraction in which compare the pixels of two different vectors and comparison of the result is shown in the form of graph. Features included in the cross correlation based classification: 1. Template matching, 2. Euclidian distances, and 3. Angel information.

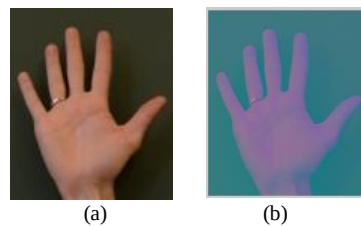
In the following subsections we describe each of the steps mentioned above.

A. Pre-Processing:

In pre-processing, basically captured image is modified in such a way that it can be easily used in the further stages of the system.

RGB To YCbCr Image Conversion: Pre-processing cannot be done on the original RGB image because it is very sensitive to the light intensity variation. So, we convert the RGB color space model into YCbCr model.

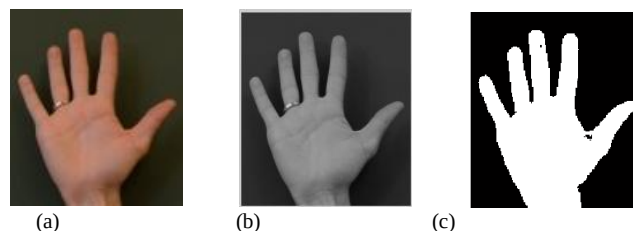
Figure 2 (a) Original image, (b) YCbCr converted image



B. Skin Color Detection:

The color information of the original image is stored into the *Cb* and *Cr* components of the resultant YCbCr matrix [17]. This information was used for skin color detection where we have put threshold on *Cb* and *Cr* components to distinguish the hand region from the entire image [12], [13]. After converting it into the YCbCr image, it was converted into the binary image. Fig. 3 shows gray scale and binary image.

Figure 3 (a) Original image, (b) Gray scale image, (c) Binary image



Noise in Binary Image: Noise in binary image can be of two types: 1. Noise in background region and 2. Noise in hand region [20]. This noise can mainly affect the performance of the algorithm. Mathematical morphology was used to remove such noise in the binary image. Morphology is a broad set of image processing operations that process images based on shapes [16]. There are two basic morphological operations which are shown below. Dilation is the process where the value of the output pixel is the maximum value of all the pixels in the input pixel's neighbourhood. So, it will assign '1' to the pixel if the neighbours are '1'. So, this operation was used to remove the noise from the hand region [17]. On the other hand, Erosion is the process where the value of the output pixel is the minimum value of all the pixels in the input pixel's neighbourhood. So, this operation was used to remove the noise from the background. After removing the noise from the binary image, we were left with the binary image having a single object without any abruption in the hand region. So that object is detected after pre-processing step.

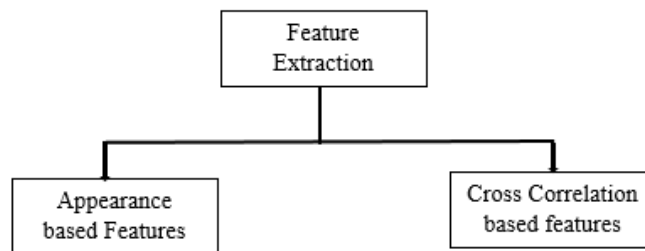
Figure 4 Binary image after pre-processing



C. Feature Extraction:

After object detection we need to recognize object by feature Extraction. In the object recognition phase, numbers of operations were performed on the binary image. Feature extraction process can be divided into two categories according to the features of image used. Fig.5 shows this classification.

Figure 5 Feature Extraction



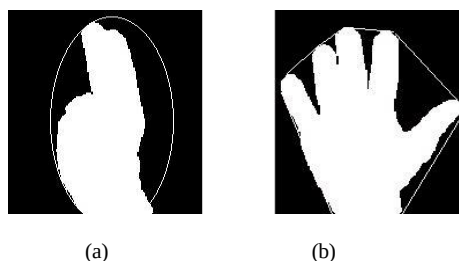
1. Appearance Based Feature Extraction:

Appearance based feature extraction include the features which involves the classification of images based on the shape, area, position of centroid, etc.

Eccentricity & Solidity: This feature taken was combination of two area based features [12], Eccentricity and Solidity. First, we have measured the eccentricity of the ellipse formed around the hand region. Eccentricity can be between zero and unity where zero eccentricity means the complete circle whereas unity eccentricity means the hand region is complete straight line. Figure 6 (a) shows the ellipse around hand region. For robustness, classification was not carried out only through eccentricity, but another feature known as *Solidity* was also obtained, and the ratio of Eccentricity and Solidity was used for the classification.

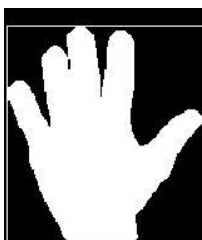
Solidity is the scalar specifying the proportion of the pixels in the *convex hull* that are also in the region [14]. So, it is basically the ratio of area of hand region and area of the convex hull. And the obtained results are used for further classification. Fig. 6 (b) shows the convex hull around hand region. And the obtained results are used for further classification.

Figure 6 (a) Ellipse around hand region, (b) Convex hull around object



Extent: It is another feature based on the areas information [12]. Extent is a ratio of the area of the hand region to the area of entire bounding box. The reason to select this feature is that it is simple to obtain and it is highly reliable as well as robust as it is the ratio. So the ratio of both the areas will remain almost same. Fig.7 shows the importance of this feature.

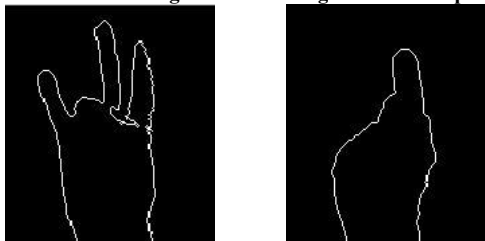
Figure 7 Bounding box across object



This symbol contains space between fingers, so its bounding box area is more as compared to the area of hand region, so its extent value would be almost unity.

Compactness Factor: The compactness measure of a shape, sometimes called the shape factor [15], is a numerical quantity representing the degree to which a shape is compact. Circle is the most compact shape. Compactness factor was determined by taking ratio of the area of the hand region to the square of the perimeter. Perimeter of the hand region was found by determining the boundary pixels. As the perimeter increases, overall compactness factor decreases. So, when the fingers are closed, the shape becomes more compact. As the results of this feature are always less than unity, we have multiplied them by 100 to classify images in simple manner. This value is not affected by the orientation and scaling the image which makes it a robust feature.

Figure 8 Perimeter for gesture "7" and gesture "1" respectively



2. Cross-Correlation Based Feature Extraction:

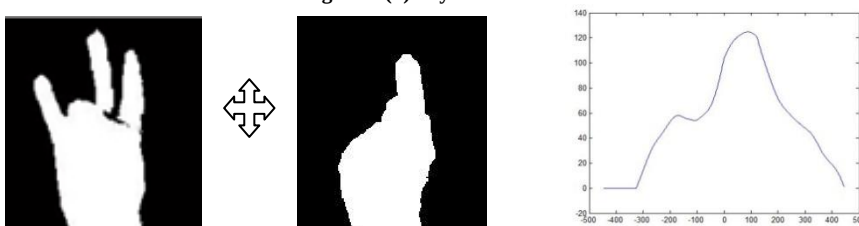
Cross correlation based feature extraction in which compare the pixels of two different vectors and comparison of the result is shown in the form of graph.

Template Matching: This feature is a promising feature which is very efficient to classify the images. In this kind of features, we have to compare a test symbol or image with every remaining image. In template matching, first step is to determine the template of the image. By *Template*, we mean the image which contains only boundary pixels. Boundary-distance vector is obtained by calculating the distance between the centroid and boundary pixels. These distances are known as *Euclidian Distances*. We have to calculate this vector for all the images, as we have to correlate them. Now cross-correlate the vectors and plot the graphs of it. After determining the graph, we have to look at the symmetry of the graph around the zero phase crossing. If they are correlated, graph will be symmetric, but if the images are different, it won't be symmetric. Fig.9 shows the symmetric and asymmetric curves.

Figure 9 (a) Symmetric curves

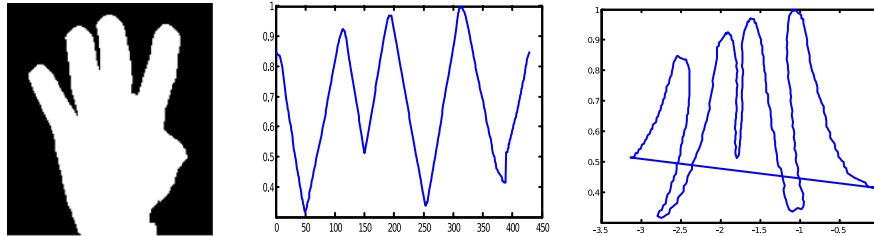


Figure 9 (b) Asymmetric curves



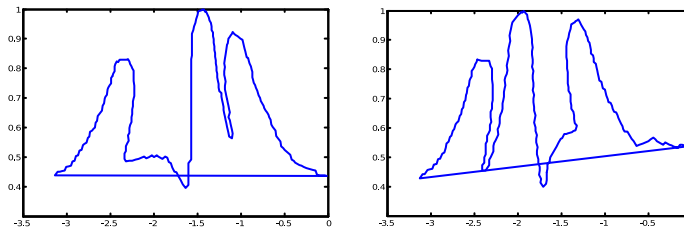
Euclidian Distances and Angle Information: We have classified remaining symbols using peak information, number of peaks in the image and the location of peaks in the image. Euclidian distances and the angle information were determined and after plotting the graphs [16], [17] all the necessary information was computed. Fig.10 shows the image with both the curves.

Figure 10 Euclidian Distances and Angle Information curves for gesture “4” respectively



Based on the position of the peak, we have distinguished some symbols. It was basically useful when number of peaks were same. It is shown in Fig.11.

Figure 11 Graph of different Peak position for gesture “7” & gesture “8” respectively



IV. EXPERIMENTAL RESULTS

The proposed algorithm was tested on two different databases. It is to note that the database images were having uniform background and all the test images were having the same dimension. Results are shown in the table 1

Table 1 Summary of results

Input Gesture	Number of Input Images	Successful Cases	Recognition rate
1	2	2	100
2	2	2	100
3	2	2	100
4	2	2	100
5	2	2	100
6	2	2	100
7	2	1	50
8	2	2	100
9	2	2	100
ALL	18	17	94.44%

Here, we find recognition rate by equation,

$$\text{Recognition Rate} = \frac{\text{No. of Recognized image}}{\text{No. of total samples of that image}} \times 100 \%$$

This table indicates that out of 18 database images, 17 were identified correctly with the success rate of around **94.44%**.

V. CONCLUSION

We proposed a fast and simple algorithm for a hand gesture recognition problem. The algorithm was developed for the hand gesture recognition system which can become useful for the communication with deaf and dump community people.

For image detection, we convert the original images to YCbCr images to avoid the shadow effect in the pre-processing stage. After that we used six possible features to classify all the images. The algorithm was tasted on the available database images and it showed 94.44% success rate. This algorithm can be implemented for real time applications.

Based on our motivating robot control application, our algorithm can be extended in a number of ways to recognize a broader set of gestures. The segmentation part of our algorithm is too simple, and would need to be improved if this method would need to be used in challenging operating conditions. We should note that the segmentation problem in a general setting is an open research problem itself. Reliable action of hand gesture recognition techniques in a general setting require dealing with occlusions, worldly tracking for recognizing dynamic gestures, as well as 3D modelling of the hand, which are still mostly beyond the current state of the art.

ACKNOWLEDGEMENTS

I would like to thank all the researchers working on this field who in one way or another guided me on achieving my goals. We would also like to express our appreciation to all other people's engr. Sweety M. Maniar, Engr. Savan J. Pindaria and specially thanks from heart to my husband Engr. Ashish A. Patel, who were kind enough to share their views with me and offered some suggestions in making this research work a success.

REFERENCES

- [1] Davis and M. Shah "Visual Gesture Recognition", IEEE Proc.-Vis. Image Signal Process., Vol. 141, No.2, April 1994.
- [2] C.-C. Chang, I.-Y. Chen, and Y.-S. Huang, "Hand Pose Recognition Using Curvature Scale Space", IEEE International Conference on Pattern Recognition, 2002.
- [3] Utsumi, T. Miyasato and F. Kishino, "Multi-Camera Hand Pose Recognition System Using Skeleton image", IEEE International Workshop on Robot and Human Communication, pp. 219-224, 1995.
- [4] Y. Aoki, S. Tanahashi, and J. Xu, "Sign Language Image Processing for Intelligent Communication by Communication Satellite", IEEE International Conf. On Acoustics, Speech, and Signal Processing, 1994.
- [5] Meenakshi Panwar and Pawan Singh Mehra, "Hand Gesture Recognition for Human Computer Interaction", in Proceedings of IEEE International Conference on Image Information Processing (ICIIP 2011), Wanknaghat, India, November 2011.
- [6] Zhong Yang, Yi Li, Weidong Chen, Yang Zheng,, "Dynamic Hand Gesture Recognition Using Hidden Markov Model", in Proceedings of 7th International Conference on Computer Science & Education (ICCSE 2012) July 14-17, 2012. Melbourne, Australia.
- [7] R. Rosales, V. Athitsos, L. Sigal, and S. Sclaroff, "3D Hand Pose Reconstruction Using Specialized Mappings", IEEE International Conf. on Computer Vision, pp. 378-385, 2001.
- [8] C. Tomasi, S. Petrov, and A. Sastry, "3D = Classification + Interpolation", IEEE International Conf. on Computer Vision, 2003.
- [9] W. T. Freeman and M. Roth, "Orientation Histograms for Hand Gesture Recognition", IEEE International Conf. on Automatic Face and Gesture Recognition, 1995.
- [10] L. Bretzner, I. Laptev, and T. Lindberg, "Hand Gesture Recognition using Multi-Scale Color Features, Hierarchical Models and Particle Filtering", IEEE International Conf. on Automatic Face and Gesture Recognition, 2002.
- [11] Ankit Multanmal Oswal, Gagan Shivarama Shetty, Mridul Anil Hiwarkar, Sanjeev Kumar Malik, Mrs. Hemangi Shinde, "Real - Time Robot Car Control Using Hand Sign Recognition", International Journal For Technological Research In Engineering Volume 1, Issue 8, April-2014.
- [12] Mohammad I. Khan, Sheikh Md. Ashan, Razib Chandra, "Recognition of Hand Gesture Recognition Using Hidden Markov Model", IEEE, pg. 150-154, 2012.
- [13] Rajat Shrivastava, "A Hidden Markov Model Based Dynamic Hand Gesture Recognition System using OpenCV", International Advance Computing Conference, IEEE, pg. 947-950, 2012.
- [14] Jaya, Ashutosh, "A Method for Hand Gesture Recognition", International Conference on Communication System and Network Technology", IEEE, pg. 919-923, 2014.
- [15] Archana, Rucha, Sanjana, Surbhi, Mithila, "Hand Gesture Recognition for Indian Sign Language", International Conference on Computer Communication and Informatics, IEEE, pg. 1-4, 2012.
- [16] Ayan, Pragya, Rohit, "Information Measure Ratio Based Real Time Approach for Hand Segmentation with a Focus on Gesture Recognition", International Conference on Intelligent Systems, Modeling and Simulation, IEEE, pg. 172-176, 2011.
- [17] Parul, Hardeep, "Neural Network Based Static Sign Gesture Recognition System", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCCE), Vol. 2, Issue 2, pg. 3066-3072, 2014.
- [18] Haris Kumar Kaura, Vipul Honarao, Sayali Patil, Pravish Shetty., "Gesture controlled Robot using Image Processing", (IJARAI) International Journal of Advance Research in Artificial Intelligence, Vol.2, 2013.
- [19] Asanterabi Malima., "A Fast Algorithm for Vision based Hand Gesture Recognition for Robot Control", IEEE International Conference on Computer Vision, 2006.
- [20] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice-Hall, 2nd edition, 2002.