

# A Comparative Analysis of Hand Tracking Algorithms for Gesture Recognition

Ambika. L. A<sup>#1</sup>

Computer Science and Engineering, Affiliated to VTU  
Belgaum, Rao Bahadur Y. Mahabaleswarappa Engineering  
College, Bellary-583104, Karnataka, India

H. Chidananda<sup>#2</sup>, Assistant professor

Computer Science and Engineering, Affiliated to VTU  
Belgaum, Rao Bahadur Y. Mahabaleswarappa Engineering  
College, Bellary-583104, Karnataka, India

**Abstract:** With the ever increasing diffusion of computers into the society, it is widely believed that present popular mode of interactions with computers will become bottleneck in the effective utilization of information flow between the computers and the human. Gesture recognition is a natural and powerful tool supporting efficient and intuitive interaction between the human and computer. The hand gesture recognition methods are robust in real world applications because they overcome the drawbacks in uncontrolled environments, including gesture hand out of the scene, complex background, skin colored regions moving in the background and face overlapping with hand. In this paper a comparative study between two tracking methods is presented. A framework for real time hand gesture recognition in uncontrolled environment is proposed which presents a robust and efficient hand tracking algorithm by wearing a black glove on hand. And also we have focused on another tracking algorithm called as free hand tracking which is based on skin color.

**Keywords:** Hand gesture recognition, tracking, complex background, segmentation.

## I. INTRODUCTION

Since their first appearance, computers have become a key element of our society. Surfing the web, typing a letter, playing a video game or storing and retrieving data are just a few of the examples involving the use of computers. And due to the constant decrease in price of personal computers, they will even more influence our everyday life in the near future. To efficiently use them, most computer applications require more and more interaction. For that reason, human-computer interaction (HCI) has been a lively field of research these last few years. To achieve natural and immersive human-computer interaction, the human hand could be used as an interface device. Hand gestures are a powerful human to- human communication channel, which forms a major part of information transfer in our everyday life. Hand gestures are an easy to use and natural way of interaction. Using hands as a device can help people communicate with computers in a more intuitive and natural way. When we interact with other people, our hand movements play an important role and the information they convey is very rich in many ways. We use our hands for pointing at a person or at an object, conveying information about space, shape and temporal characteristics. We constantly use our hands to interact with objects: move them, modify them, and transform them. In the same

unconscious way, we gesticulate while speaking to communicate ideas ('stop', 'come closer', 'no', etc). Hand movements are thus a mean of non-verbal communication, ranging from simple actions (pointing at objects for example) to more complex ones (such as expressing feelings or communicating with others). In this sense, gestures are not only an ornament of spoken language, but are essential components of the language generation process itself [1].

Gesture recognition is the process by which the *gestures* made by the user are *recognized* by the receiver. Gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head, face, or body with the intent of: 1) conveying meaningful information or 2) interacting with the environment [2]. Normally gesture recognition systems are divided into three stages namely image processing followed by tracking succeeded finally by the recognition stage as depicted in Figure 1. Generally, there exist many-to-one mappings from concepts to gestures and vice versa. Hence, gestures are *ambiguous* and *incompletely specified*.

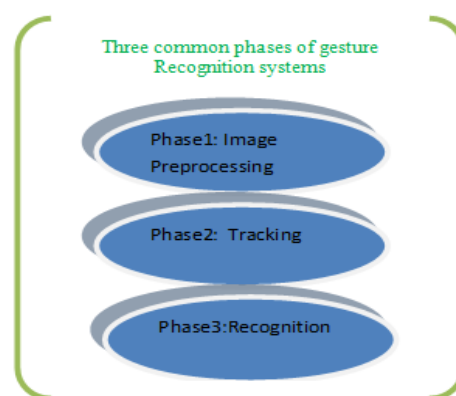


Fig. 1 A gesture recognition approach

Gesture recognition is a topic pursued with the goal of interpreting human motions via mathematical algorithms. Gesture recognition can be seen as a way for a computer to begin to understand human body language, thus building a richer bridge between machines and humans than primitive text

interfaces or even graphical user interfaces (GUI), which still limit the majority of input to keyboard and mouse. Hence, gesture recognition is a process by which the system will know, what is going to be performed by the gesturer. Gesture recognition can be conducted with techniques from computer vision and image processing. Gesture recognition enables humans to interface with the machine and interact naturally without any mechanical devices [3]. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that the cursor will move accordingly. This could potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant. This paper is organized as follows: Section II gives basic theoretical considerations. Some comparative study on segmentation is presented in literature survey of Section III. Section VI gives problem statement. Section V presents the system design for the proposed gesture recognition system. Some of the observations are made in Section VI. The effort ends in Conclusion that is in Section VII.

## II. BASIC THEORITICAL CONSIDERATIONS

The block diagram for gesture recognition system is shown in the figure 2

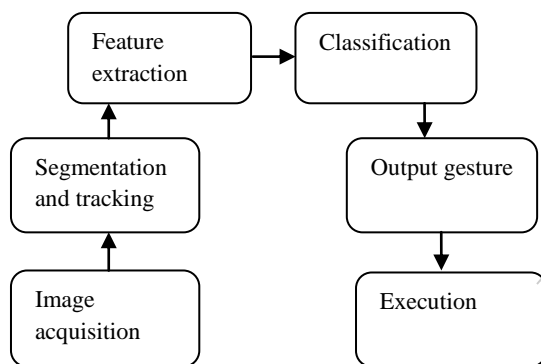


Fig. 2: Block diagram for gesture recognition system

Image acquisition is the first process in which a set of image frames are captured by using low cost web camera. This stage involves preprocessing such as scaling. Hand segmentation is necessary to track the movement of hand. Segmentation partitions the image into its constituent parts or objects. Once the hand is tracked there is a need to extract the most important feature points from the available data points. The next step is a feature extraction; it is a special form of dimensionality reduction. If the input data to an algorithm is too large to be processed and if it is redundant then it is transformed into reduced set of features. Transformed input data into set of features is called as feature extraction. Once the features are extracted, classifier plays a vital role in recognition process. Classifier takes feature set as input and gives a class labeled output, which are the required gestures. Each class is mapped to particular functions and executing actions related to gesture.

The following figure3 and 4 shows the conversion from RGB to YCbCr and HSI respectively.

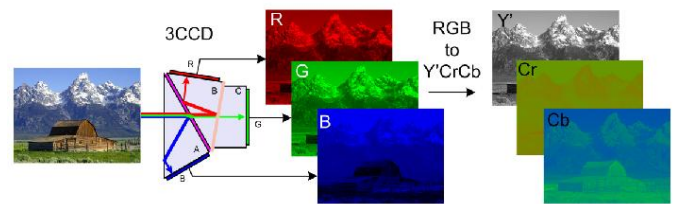


Fig. 3: RGB to YCbCr conversion

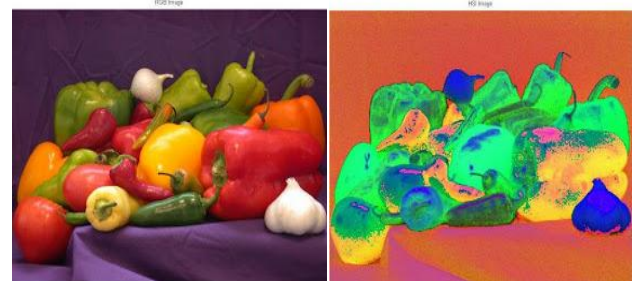


Fig. 4: RGB to HSI conversion

## III. PREVIOUS WORK

To improve the interaction in qualitative terms in dynamic environment it is desired that the means of interaction should be as ordinary and as natural as possible. The gestures performed by users must be logically explainable for designing a good human computer interaction [2]. The current technologies for gesture recognition are not in a state of providing acceptable solutions. One of the major challenges is complexity and robustness associated with the analysis and evaluation for gesture recognition. Dynamic conditional random field (DCRF) model is used for foreground object and moving shadow segmentation in indoor video scenes. An efficient approximate filtering algorithm is derived for DCRF model to recursively estimate the segmentation field from the history of observed images. This approach can accurately detect moving objects even in grayscale video sequences [4]. Computer vision face tracking is an active and developing field, yet the face trackers that have been developed are not sufficient for our needs. We want a tracker that will track a given face in the presence of noise, other faces and hand movements. To track colored objects in video frame sequences the color image data has to be represented as a probability distribution. Color distributions derived from video image sequences change over time, so the mean shift algorithm has to be modified to adapt dynamically to the probability distribution that it is tracking. CAMSHIFT (continuously adaptive mean shift) algorithm is a simple, efficient colored object tracker [5]. Proper hand segmentation from the background and the other body parts of the video is the primary requirement for the design of a hand gesture based application. It presents a robust and efficient hand tracking as well as segmentation algorithm which efficiently handles problems of varying lighting conditions [6]. Multi-stage classification procedure reduces the processing time substantially while achieving almost the same accuracy as compared to a much slower and more complex single stage classifier. In cascade of classifier, at each stage a

classifier is trained to detect almost all objects of interest while rejecting a certain fraction of non-object patterns [7]. The major factor that disturbs automatic gesture recognition is illumination change. It constructs the background model based on hue and the hue-gradient and then robustly extracts the object contours. When tested in complex environment, if hand region is similar to background then the hands contour was lost [8].

#### IV. PROBLEM STATEMENT

When gesture recognition system tested in complex environment some of the hand contour parts are lost. To overcome this issue we are making use of robust and efficient algorithm. In addition, problems such as skin color detection, complex background removal and variable lighting condition are found to be efficiently handled by the system. Due to dynamic background some noise will be present in the segmented image so it will be removed with the help of this adaptive technique.

#### V. DESIGN OF PROPOSED SYSTEM

Here we employ MATLAB software with a low cost web camera, which reduces the computational complexity so that the techniques can operate in real time. Tracking is the process of locating and normalizing the hand. There are several methods that can be used for tracking, each with different success and feature conditions. Two different types of hand tracking algorithms are described in this paper. One algorithm is developed by free hand based on skin color and the other by using black glove. Both of them have advantage as well as disadvantage over each other. These two algorithms are described one by one as given below:

##### A. Hand tracking based on skin color:

Hand segmentation is necessary to track the movement of the hand. Here tracking is based on the skin color. In this algorithm skin color segmentation method is used only to extract the skin regions from the video frame [5]. By using this background elements are eliminated and only face and hand regions are extracted because they have almost same color. After this, the face region is removed from the frame by using face detection algorithm. As we are interested only on the hand region face is initially removed. In order to separate the image intensity from the color information, the original color plane is converted to two separate HSV and YCbCr color planes. Morphological processing is done over the image planes in order to extract the image components that are useful in representation. Firstly image planes are converted to binary form to apply some morphological operation like erosion, dilation etc., to extract the skin regions. After that logical 'AND' operation is done between to get the most probable skin region i.e., hand. Once we get the skin color segmented regions, the largest connected segment corresponds to the palm region of the hand. Then the centroid is determined for the palm region. Figure 5 shows flow chart of this algorithm.

The algorithm can be summarized as below:-

Step 1- Face region is detected and subtracted from the input image frame.

Step 2- The RGB frame is converted into two color planes, i.e. HSV and YCbCr.

Step 3- Both image frames is converted to binary plane for doing some morphological operation to extract the skin regions.

Step 4- A logical 'AND' operation is performed between the planes, which gives segmented hand region.

Step 5- Centroids of segmented region is calculated using moments calculation.

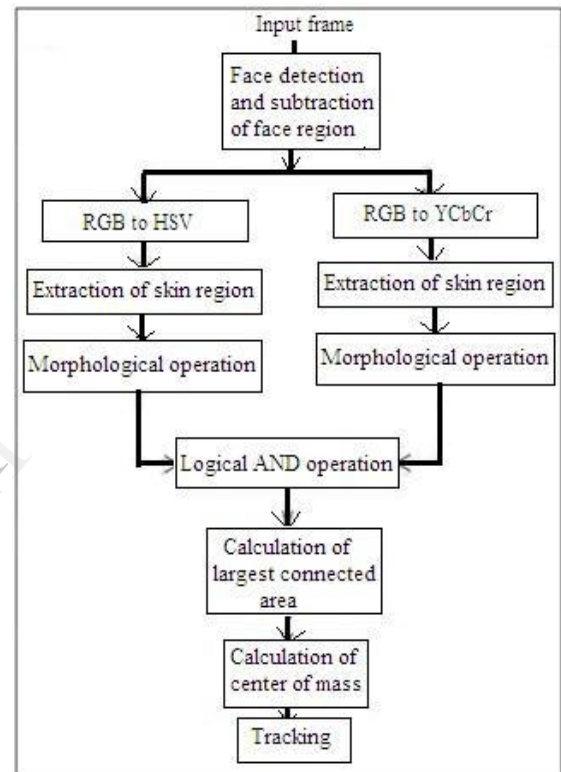


Fig. 5: Hand tracking based on skin color

##### B. Hand tracking by using black gloves:

In this algorithm the palm has a black glove. HSV color plane is the most suitable color plane for color based image segmentation [9]. So, we have converted the color space of original image frame of the camera, i.e. from RGB to HSV plane. By setting the hue value for the hand gloves, we can easily eliminate all the other parts of the image frames. Finally, this output is converted to binary form and certain morphological operation likes erosion, dilation etc. is carried out. It results in a noiseless segmented hand or region of interest (ROI) that can be used subsequently. Then, the centroid of the segmented hand is determined and subsequently the centroid is used for making gesture trajectories. Figure 6 shows flow chart of this algorithm.

The algorithm can be summarized as below:-

Step 1- The input image frame is first converted to HSV color plane from RGB plane.

Step 2- Hue value is then set properly for the color of the hand glove.

Step 3- HSV plane is converted to binary plane with proper threshold.

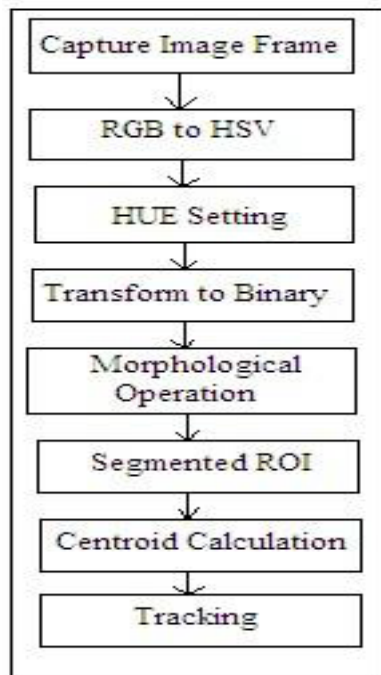


Fig .6: Hand tracking by using black gloves

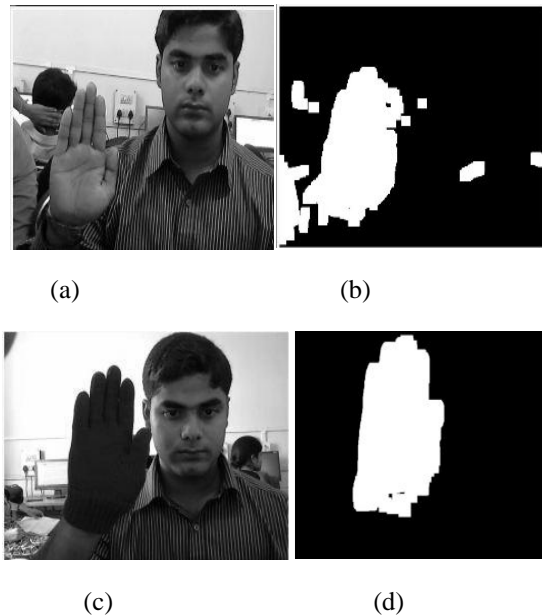
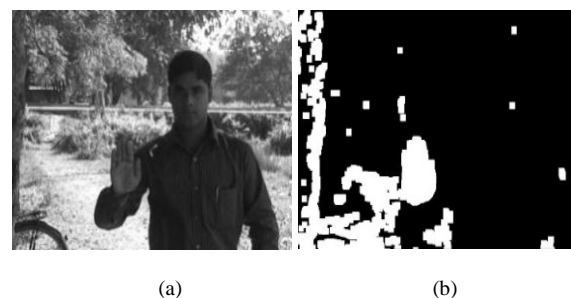


Fig 7: Segmentation result at static background condition

Figure 7a is the input to free hand tracking technique at static background with poor light condition, which gives a noisy output as shown in Figure7b. But the same samples when are applied as input as shown in Figure7c, to gloved hand tracking technique gives a better segmented output as shown by Figure7d. We have performed a number of trials by changing the lighting conditions, but this technique is robust to all these variable conditions and gave the same output. This is because this system shows no effect to the hue value of the glove. Again, Figure8 shows some of the test results at dynamic background conditions. To know the robustness of the system, various environments are considered. Figure8b is the output from free hand tracking technique for the input as shown in Figure8a. This gives a noisy segmented result. Figure8c and 8d shows the input and output gloved hand tracking technique respectively. Hence we can conclude that this technique is robust to lighting and complex background conditions and it does not depend on any situation.



### C .Calculation of centroids:

We have measured the centroids of the segmented hand region by a simple way, which is same for both the algorithms. The centroid  $(x, y)$  of the segmented hand region  $R$  is arithmetic mean of the co-ordinates in the  $x$  and  $y$  directions, as given in equation (1).

$$x = \frac{1}{|R|} \sum_{(u,v) \in R} u \quad \text{and} \quad y = \frac{1}{|R|} \sum_{(u,v) \in R} v \quad (1)$$

(1)Equation to calculate centroids

## VI. OBSERVATION

A few samples are recorded by a webcam and indoor-outdoor sets are created for use with these algorithms. In this paper we have included the segmented output result found from the algorithms. The samples consider illumination and background variation.

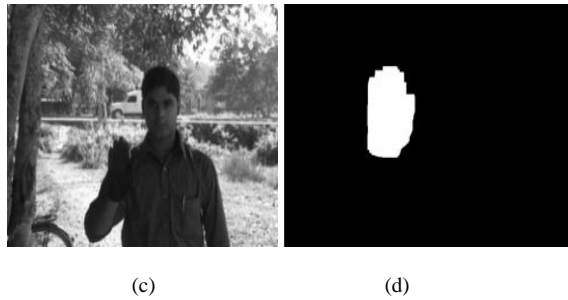


Fig 8: Segmentation result at dynamic background condition

Table I: Comparison of segmented output of the algorithms

Segmentation models	Normal	Complex background	Dynamic lighting	Dynamic background
Hand tracking based on skin color	Working	Working	Working	Noisy output
Hand tracking using black gloves	Working	Robust	Robust	Robust

After recognizing gestures, the system uses these different hand gestures to browse images in image browser. Figure 9 shows some hand gestures along with their assigned commands for browsing the images in image browser

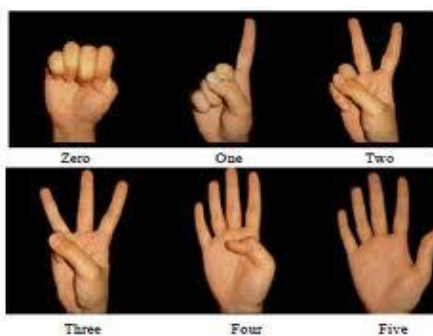


Fig 9: Hand gestures

Though the gestures have been mapped for commands used in image browsing in Figure 10, the same gesture vocabulary could be reused for mapping different set of commands according to different range of applications like controlling games, power point presentations etc. This makes the gesture recognition system more generalized and adaptive towards human computer interaction.

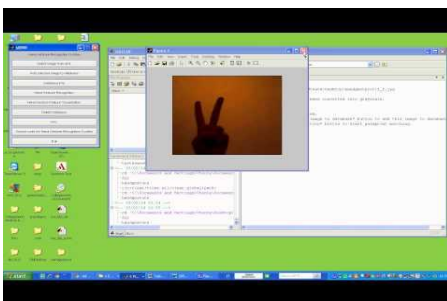


Fig 10: Browsing images using hand gestures

## VII. CONCLUSION

Hand tracking based on skin color algorithm gives noisy output in varying lighting conditions and dynamic background situation hence it is very sensitive when compared to black gloved hand tracking algorithm. Apart from these drawbacks, there are few limitations such as: gesturer should always wear full sleeve to easily detect the palm part from the rest of body parts. When more than one gesturers are there then it is difficult to detect the correct gesturer. When the hand comes in front of the face then it is difficult to separate it from the face. Hand tracking by using black gloves method is robust to all variable conditions and gives the same output hence it is not dependent on any background and lighting variations. Real time recognition process with no time delay is one of the features offered by this algorithm. We can use it for real time purpose in any environment.

## ACKNOWLEDGEMENT

I Ambika L.A would like to thank my guide Mr. H.Chidananda, Assistant professor who supported me in preparing this paper.

## REFERENCES

- [1] Siddharth S. Rautaray, Anupam Agrawal, "Real Time Multiple Hand Gesture Recognition System for Human Computer Interaction", IJ. Intelligent Systems and Applications, 2012, 5, 56-64.
- [2] S. Mitra, and T. Acharya, "Gesture Recognition: A survey," *IEEE Transactions on Systems, Man and Cybernetics (SMC) - Part C: Applications and Reviews*, vol. 37(3), pp. 211-324, 2007.
- [3] E. S. Nielsen, L. A. Canalis and M. H. Tejera, "Hand Gesture Recognition for Human-Machine Interaction, *Journal of WSCG*, vol. 12, no. 1-3, 2004.
- [4] Y. Wang, K. F. Loe and J. K. Wu, "A Dynamic Conditional Random Field Model for Foreground and Shadow Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 279-286, 2006.
- [5] G. R. Bradski, S. Clara, "Computer Vision Face Tracking For Use in a Perceptual User Interface", *Intel Technology Journal* Q2'98, 1998.
- [6] Dharani Mazumdar, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, "Gloved and Free Hand Tracking based Hand Gesture Recognition", *ICETACS 2013*.
- [7] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," *In Proceedings of ICIP02*, pp. 900-903, 2002.
- [8] Yoo-Joo Choi, Je-Sung Lee and We-Duke Cho, "A Robust Hand Recognition In Varying Illumination," *Advances in Human Computer Interaction*, Shane Pinder (Ed.), 2006.
- [9] M. H. Yang, N. Ahuja, M. Tabb, "Extraction of 2D Motion Trajectories and its Application to Hand Gesture Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1061-1074, 2002.