A Static Hand Gesture Recognition System to Recognize the Total number of Fingers

D.K. Vishwakarma* Sahib Majithia† Nikhil Kumar Mishra‡

Department of Electronics and Communication Engineering,

Delhi Technological University

New Delhi, India

Email: *dvishwakarma@gmail.com ,†sahibmajithia@gmail.com, ‡nkmishra1997@gmail.com

Abstract—**Gesture recognition aims at recognizing meaningful expressions of motion by a human, involving the hands, arms, face, head, or body. It is of great importance in designing an intelligent and efficient human-computer interface. We take into account the two possible fixed geometries to work on. The proposed methods follow the procedure of Preprocessing: which helps with noise removal and image enhancement; Segmentation of hand region: uses skin likelihood method to extract skin color; Feature extraction: Morphological and geometry based functions are used to extract the fingers; and active fingers are counted by method of Rule based Classification. In order to test performance, an experiment is conducted using standard and a self-generated dataset of images. The accuracy achieved on these dataset is greater than on similar state of arts.**

Keywords: *Hand Gestures, Skin Segmentation, Active Fingers, Finger Count, Mask Generation, Recognition, Low Resolution*

I. INTRODUCTION

In the present day framework of interactive, intelligent computing, an efficient humancomputer interaction is assuming utmost importance in our daily lives. Gesture recognition can be termed as an approach in this direction. Gesture recognition is the mathematical interpretation of a human motion by a computing device

There are methods which used additional hardware devices such as data glove devices and color markers to easily extract a comprehensive description of gesture features for recognition [1]. Neural Network classifier has been applied for gestures classification [2] but it is time consuming and when the number of training data increases, the time needed for classification are increased too .Orientation histogram method [3] has some problems which are; different gestures could have similar orientation histograms, and similar gestures might have different orientation histograms. Other methods based on the appearance of the hand use skin color to segment the hand and extract necessary features, these methods are considered easy, natural and less computationally costly compared with methods mentioned before [1]. This paper proposes two methods for counting the number of active fingers. These are proposed based on the geometric facts of the hand region.

Since the proposed method works on geometric properties and does not require any extracted features; it works well on low-resolution images which is a great advantage for developers regarding data storage.

The algorithm first helps with skin segmentation from the region of interest. Further, the number of active fingers are counted after operating them with a range of morphological operations. In this paper section 2 and 3 describe the algorithms proposed. Further sections discuss the Results (4), Conclusions (5) and Future scope (6) of our work.

II. HAND SEGMENTATION

Skin segmentation obtains the hand region. The skin segmentation has been applied correctly to Region of Interest of hand region. This does not cause susceptibility as the Region of Interest can be easily extracted by different depth estimation strategies[4], which include:

Active methods: Light-based depth estimation/Pattern projection/Ultrasounds based methods

Passive methods: Monocular depth estimation/Image structure/Points tracking or Optical flow/ Multiview solutions for the depth estimation

The Region of interest can also be extracted by using instruments like Kinect which uses an RGB camera with a depth sensor.

Fig. 1. GESTURES USED WITH EXTRACTED ROI

A. SKIN SEGMENTATION

The most important part of hand gesture recognition is the skin segmentation. Segmentation accuracy determines the eventual success or failure of an automated analysis procedure.For segmentation, we first convert the RGB colored image into the YCbCr color model.

The transformation of the original RGB colour image is done using the following equations:

$Y = 0.2990R + 0.5870G + 0.1140B$

$Cb = -0.1687R - 0.3313G + 0.5000B$

$Cr = 0.5000R - 0.4187G - 0.0813B$

The mean values of chrominance components are used to extract the skin color of the hand by dropping the luminance component because the variation in skin color is more in intensity than in chrominance. In YCbCr color space, the components Cb and Cr represent the chrominance and component Y accounts for the luminance.

In the earlier works, the YCbCr color model has been used by a basic mathematic operation of thresholding on chromaticity and luminance values. No method has yet been settled which determines the thresholding accurately.So, alternatively, we use a modified approach called a skin likelihood model for skin segmentation.

This model is designed from the maximum likelihood criterion, which has the parameters: mean vector and covariance matrix that maximizes the likelihood function. The Gaussian likelihood function[5] has single maxima, and the parameters mean vector (μ) and covariance matrix (Σ) can be computed as:

$$\mu_i = \frac{1}{n} \sum_{k=1}^{\infty} x_k \tag{1}$$

Where i $\in$ {Cb,Cr} Cb and Cr are the values of chrominance in YCbCr color space.

$$\Sigma = \frac{1}{n} \sum_{k=1}^{\infty} (x_k - \mu)(x_k - \mu)^T \tag{2}$$

The probability distribution of skin class can be written as:

$$p(\frac{x}{skin}) = g(x, \mu, \Sigma) \tag{3}$$

$$g(x, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} (det(\Sigma))^{\frac{1}{2}}} \times exp(-\frac{1}{2}\delta^2) \tag{4}$$

Where d is the dimension of the feature vector, and $\delta^2 = (x - \mu_i)\Sigma^{-1}(x - \mu_i)^T$ .From equation 4, the pixels can be classified as skin and non-skin color pixels by using the decision boundary. The probability density distribution of the formed Gaussian model is represented in figure: Plotting
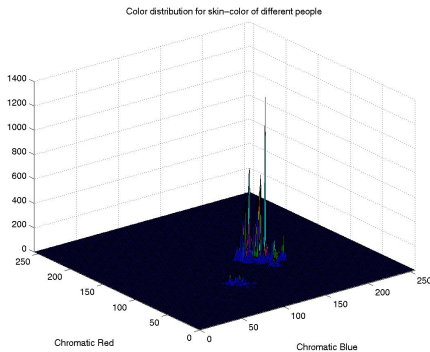

Fig. 2. Graph describing variation of skin tones of human race in Chromatic colour model

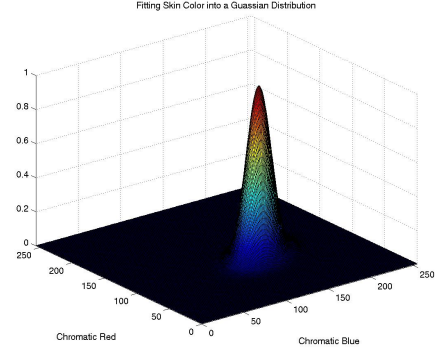a Gaussian Curve for this variation yields the Graph in figure 3.

Fig. 3. Gaussian curve of variations in Fig.1

This Method has a great accuracy and calibration advantage over the YCbCr model and is thus more user-friendly. It can be calibrated for all fairer to darker skins with just a change in input skin section (which was trained for the Gaussian); in contrast to change in Cb and Cr values which demand a technical blueprint.

B. PSEUDO CODE:

-imresize([200 200])
SKIN SEGMENTATION:
$trainskinmodel :$
$function\_outputs[mean(cr), mean(cb), cov(cr, cb)]$
$x = [(cr - rmean); (cb - bmean)];$
$likely\_skin(i, j) = [power(2 * pi * power(det(rbcov), 0.5), -1)] * exp(-0.5 * x' * inv(rbcov) * x);$
For every pixel we calculate :
$likely\_skin(i, j) = [power(2 * pi * power(det(rbcov), 0.5), -1)] * exp(-0.5 * x' * inv(rbcov) * x);$
%This gives the Gaussian of the image.
$BW = im2bw(likely_skin, thresh);$ %Morphological operations for segmenting the hand shape accurately
$imclose; imdilate; imfill('holes');$
Find the hand(palm width)
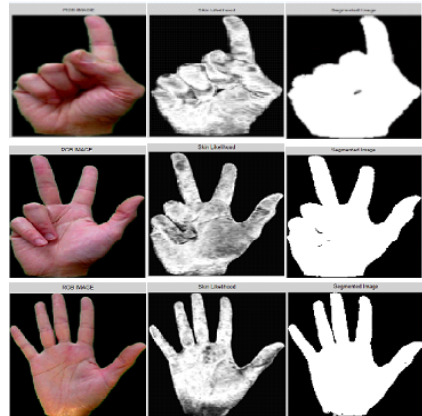
III. PROPOSED METHODS

A. Method 1: MASKING APPROACH

After acquiring a binary image of the segmented skin region, we extract the count of active fingers by a distinct system of morphological operations.

1) MORPHOLOGY: Morphology[6] is an extensive collection of image processing procedures that process images based on shapes.

Dilation and Erosion work (at least conceptually) by translating the structuring element to several points in the input image, and analyzing the overlap between the translated kernel coordinates and the input image coordinates. For example, in the case of erosion, the output coordinate set consists of just those points to which the origin of the structuring element can be translated, while the element still remains entirely 'within' the input image.

Mathematically, the main morphological operations we use are denoted as follows:

  a) EROSION:

$$(A \ominus B) = [z|(B)_z \subseteq A]$$

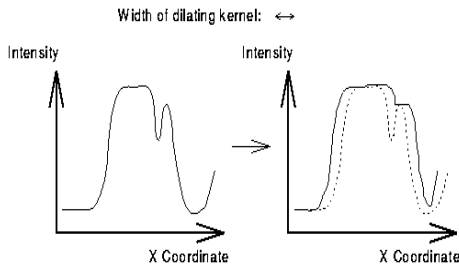Width of eroding kernel:
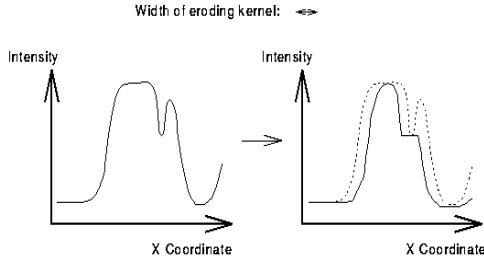
  b) DILATION:

$$(A \oplus B) = \left[z|(\widehat{B})_z \cap A \neq \phi \right]$$

imopen:

$$(A \ o \ B) = [(A \ominus B) \oplus B]$$

imclose:

$$(A \cdot B) = [(A \oplus B) \ominus B]$$

Width of dilating kernel:

2) OPERATIONS: The palm width calculation is needed for obtaining the finger width[7].

The finger width lies between 1/3 to 1/4 times the palm width depending on the stretch of fingers.

An erosion filter is then designed which is of the dimension of the finger girth. This filter is employed for erosion on the binary segmented image. Dilation of this image helps obtain the mask (here palm).

Morphological opening, closing, and other fitting operations should be executed to accomplish a noise-free and an explicitly shaped mask.

The subtraction of the mask[8] from the binary image provides us the active fingers of the image. Further thresholding for the area of the acquired portions presents us with the precise count of active fingers.
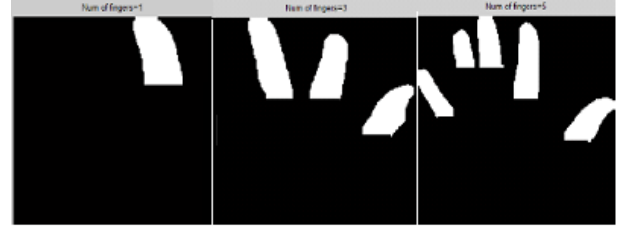
3) PSEUDO CODE: METHOD 1

3 <ratio <4

$erode\_filter = ones(1, floor(width/ratio)$; % remove small components such as noises and uneroded finger parts

$bwareaopen(eroded_image, 40)$; %dilate to obtain the original palm shape and size.

$fingers = palm. * BW$

$regionprops \rightarrow$ Area threshholding to count the no. of bounding box to count the number of fingers.

B. Method 2: GRADIENT APPROACH

After acquiring a binary image of the segmented skin region, we extract the count of active fingers by using the information of change of gradients at the finger-tip region. Once the centroid of the binary image is obtained, the shortest distance of the boundary of the silhouette from this centroid point is determined. This minimum distance is determined in all directions from the centroid and the area thus formed depicts the mask for the binary silhouette. This area usually comprises the palm region of the hand and hence forms the base of the mask[9]. Mask subtraction then gives us the fingertip section in addition to the wrist area.

Gradients are calculated by using filters :

Gx: forward horizontal difference

Gy: forward vertical difference

$$Gradient = \sqrt{G_x^2 + G_y^2}$$

We use sobel gradient operator for our method.

Gradient map of the obtained image is calculated. The geometry of the hand infers that the wrist area has a small change of slope at the boundary region though the change is high for the fingertips.
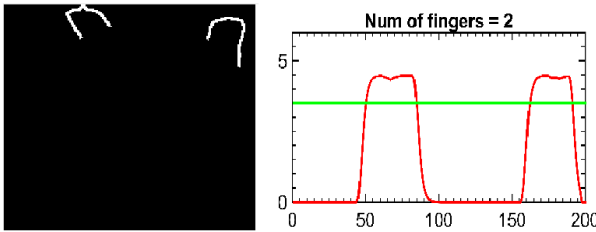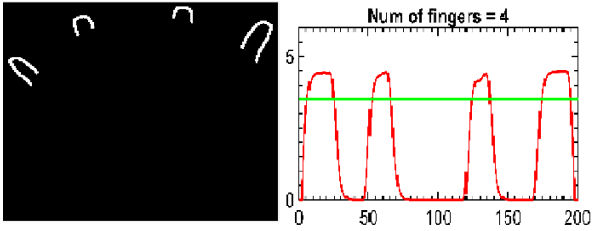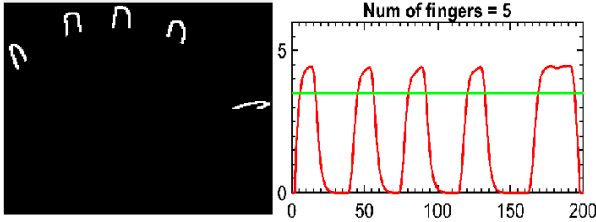
-1	0	+1
-2	0	+2
-1	0	+1

TABLE I
Gx

+1	+2	+1
0	0	0
-1	-2	-1

TABLE II
Gy

IV. RESULT

The rule-based classifiers learned model is represented as a set of if-then rules. Let the correct number of tests be c out of a total of t tests. Accuracy is then given by $(c/t) * 100$.

A. STANDARD DATA-SET

Thus a suitable threshold is established to segregate the wrist area.

The gradient map is passed through a low-pass filter to remove the noise effects. The low-pass filter also helps identify the false high-gradient-change locations.

The final gradient map is a plot for column locations, and thresholding identifies the count of fingers.

We applied our algorithms on standard Massey University database [10] of images available online.We used only 5 hand gestures from the data-set

Accuracy by Method 1: - 93.622

Accuracy by Method 2: - 90.136

TABLE III
CONFUSION MATRIX OF METHOD 1

Finger Count	1	2	3	4	5
1	92.75%	7.25%	0%	0%	0%
2	0%	100%	0%	0%	0%
3	0%	0%	92%	8%	0%
4	0%	0%	1.4%	97.2%	1.4%
5	0%	0%	0%	9.23%	86.16%

TABLE IV
CONFUSION MATRIX OF METHOD 2

Finger Count	1	2	3	4	5
1	88.4%	11.6%	0%	0%	0%
2	1.43%	94.29%	4.28%	0%	0%
3	1.33%	12%	85.34%	1.33%	0%
4	1.4%	1.4%	7%	88.8%	1.4%
5	0%	0%	0%	6.15%	93.85%

B. OUR DATA-SET

1) PSEUDO CODE: METHOD 2

$erode\_filter = strel('disk', 5);$ % remove small components such as noises and uneroded finger parts

$[Gmag, Gdir] = imgradient(erodedimage);$

Find the change and direction of gradient at the boundary of the entire hand region.

$pdist$(co-ordinates of each boundary point and centroid ,'euclidean');

Find the Euclidean Distance between every boundary point from the center.
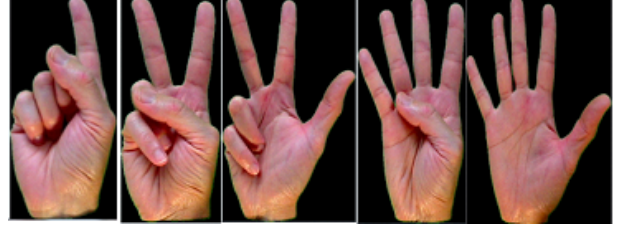
Low pass filter → Removes the noise effects, identifies the false high-gradient-change locations.

Plot the final gradient map → column locations, and thresholding identifies the count of fingers.

To have an understanding of real time results, we collect our own database of images and further test the accuracy. We collected hand gestures of different Indian persons.

Accuracy by Method 1 : - 96
Accuracy by Method 2 : - 98

TABLE V
CONFUSION MATRIX OF METHOD 1

Finger Count	1	2	3	4	5
1	100%	0%	0%	0%	0%
2	0%	90%	10%	0%	0%
3	0%	0%	100%	0%	0%
4	0%	0%	0%	100%	0%
5	0%	0%	0%	30%	70%

TABLE VI
CONFUSION MATRIX OF METHOD 2

Finger Count	1	2	3	4	5
1	100%	0%	0%	0%	0%
2	0%	100%	0%	0%	0%
3	0%	0%	100%	0%	0%
4	0%	0%	0%	100%	0%
5	0%	0%	0%	30%	70%

TABLE VII
COMPARISON OF VARIOUS METHODS

Method for feature extraction + classifier	Accuracy
Bayes classifier[11]	83.40%
PCA	72.73%
Kumar's Autoencoders[12]	83.36%
Our Method-1	93.62%
Our Method-2	90.13%

V. CONCLUSION

This paper deals with two algorithms in order to recognize the hand gestures and also compares their efficiencies with each other as well as established methods[13]. In our work, the geometry of the hand has been used for the recognition of hand gesture. It is a traditional type of algorithm and is thus robust.The proposed methods do not require any training phase[14] to identify the hand gestures so is better than neural networks. Sign Language Recognition, Robot Control automation of home appliances and presentation control are few of many applications of hand gesture recognition systems. Usage of hand gestures can be extended to control real time applications like VLC media player, paint, pdf reader etc. This wireless control of embedded instruments is a great progress towards the field of embedded vision and automation.[?]

REFERENCES

[1] Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi, Akira Iwata, (2002)"A Rotation Invariant Approach On Static-Gesture Recognition Using Boundary Histograms And Neural Networks", IEEE Proceedings of the 9th International Conference on Neural Information Processing, Singapura.
[2] Kouichi M., Hitomi T. (1999) "Gesture Recognition using Recurrent Neural Networks", ACM conference on Human factors in computing systems: Reaching through technology (CHI '91), pp. 237-242. doi: 10.1145/108844.108900
[3] W. T. Freeman and Michal R., (1995) "Orientation Histograms for Hand Gesture Recognition", IEEE International Workshop on Automatic Face and Gesture Recognition
[4] Pablo Revuelta Sanz, Beln Ruiz Mezcua and Jos M. Snchez Pena" Depth Estimation An Introduction"
[5] D.K. Vishwakarma and Rajiv Kapoor, "An Efficient Interpretation of Hand Gestures to Control Smart Interactive Television", International Journal of Computational Vision and Robotics, July, 2015,Impact Factor: 0.22, (In Press), (Pub.: Inderscience, UK).
[6] http://homepages.inf.ed.ac.uk/rbf/HIPR2/matmorph.htm
[7] Ram Rajesh J, Nagarjunan D, Arunachalam RM and Aarthi R, "Distance Transform Based Hand Gestures Recognition For Powerpoint Presentation Navigation ", Advanced Computing: An International Journal (ACIJ), Vol.3, No.3, May 2012
[8] D.K. Vishwakarma; Maheshwari, Rockey ; Kapoor, Rajiv, "An Efficient Approach for the Recognition of Hand Gestures from Very Low Resolution Images",in 5th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Gwalior, MP, pp.467-471, April 2015. DOI: 10.1109/CSNT.2015.84.
[9] D.K. Vishwakarma and Rajiv Kapoor, "Simple and Intelligent system to recognize the expression of speech disabled person", 4th IEEE international conference on Intelligent Human Computer Interaction, pp: 1-6, 27th to 29th December, 2012, Kharagpur India, DOI:10.1109/IHCI.2012.6481804.
[10] A.L.C. Barczak, N.H. Reyes, M. Abastillas, A. Piccio and T. Susnjak IIMS,Massey University, Auckland, New Zealand "A New 2D Static Hand Gesture Colour Image Dataset for ASL Gestures", Res. Lett. Inf. Math. Sci., 2011, Vol. 15, pp. 12?20
[11] Marimpis Avraams "Static Gesture Recognition Combining Graph and Appearance Features", International Journal of Advanced Research in Artificial Intelligence,Vol. 3, No. 2, 2014
[12] Kumar, Varun, Nandi, G. C. and Rahul, K "Static Hand Gesture Recognition Using Stacked Denoising Sparse Autoencoders ", IEEE International Conference on Contemporary Computing, 2014.
[13] Rafiqul Zaman Khan and Noor Adnan Ibraheem, "Hand Gesture Recognition: A Literature review", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012
[14] D.K. Vishwakarma, Pridarshani and Kuldeep Singh, "A framework for the recognition of hand gesture in static postures",in IEEE Conf. on Computing Communication and Automation, Greater Noida, UP, 2016, pp. 294-298. DOI:10.1109/CCAA.2016.7813732.