International Journal of Research
Available at
https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 09
September 2020

Adaptive Skin Colour Segmentation For Sign Language Recognition

[1]Pankaj Prajapati,[2]Alok Mishra

***Abstract*—In this paper sign language recognition system is implemented via hand gesture recognition system. For detecting hand it is very important to separate out the hand in various light condition and in cluttered background. Hand gesture can be captured by any simple camera. So RGB to YCbCr conversion is used for an adaptive skin colour segmentation. The ability of this adaptive skin colour segmentation is it can segment the out the skin region so well with very poor light conditions and whatever background except skin colour background. After that Median filter is used to remove the background noise and smoothen the edges of hand. Then Hu's Moment is calculated for extracting necessary feature. Database of hand gestures is created for 26 alphabets and trained using Artificial Neural Network(ANN).The accuracy of the system is better than previous work because of adaptive method for skin colour segmentation. Feature scope of this work is real time hand gesture recognition for sign language recognition.**

Keywords*—sign language recognition; RGB; YCbCr; adaptive skin colour segmentation, artificial neural network, median filter, Hu's moment Invariant.

I. Introduction

Gesture play an important role for communication purpose. For deaf and dumb people it is very important way to communicate. So to make them educated there is a need of computer so that physically disabled people can learn computer and also they can learn sign language. There are different types of sign language in the world for e.g. Indian sign language (ISL) and American Sign Language etc. So through human computer interaction for hand gesture recognition they can learn the sign language online.

Hand gesture recognition system is also helpful in a home automation and gaming purpose. Through the hand gesture recognition it is possible to control the relay. A vast research is going on for human computer interaction (HCI) for its usefulness.

Gesture recognition typically includes two tactic. One is static hand gesture recognition [5] and yet another is to evaluate active hand movements during this method hand movement is tracked and recognized in real time [1] [4].

There are different subsequent methods in which, the framework is apart from the earlier system from the different performed ways.

- A high accuracy with different hand gestures.
- A low processing power is required because of highly efficient computational operation of artificial neural network.
- Light variation causes very negligible impact on our system.
- Colour segmentation of Adaptive skin.

In this paper author try to give a solution methodology for vision based hand gesture recognition for sign language recognition. The feature of vision based system is one would not has to use any kind of glove for hand segmentation process. Hand segmentation is predicated on skin colour detection [5] [7]. RGB image is converted into YCbCr image. Y component signalize luminance of the pictorial image but Cb and Cr inhere chrominance of the pictorial image.

II. Related Work

There are customarily two types of category for image based hand gesture recognition. One is three dimensional based hand exemplary and another is two dimensional based style [2] [3] or it is appearance based. The three-dimensional hand geature recognition technique isimplemented using mocrosoft discovered kinect camera which also calculates the deepness of the image.

Appearance-based exemplarymethod consits necessary feature is extracted from the segmented hand image and these featured is compared with trained database in which feature had already exatracted. Through that Geature is recognized. This is approach of this system because of its cost and techonology effectiveness.

In [1], To separate out the hand images RGB to YCbCr conversion is used. Threshoding is done by setting the Cb and Cr values. Hu's Mpment is find out which is invariant to it's rotation, translation and scaling.

In paper [2] for recognizing hand gesture, hand detection and subtraction of faces is done. Differentiating with different hand gestures counter comparison algorithm is used. Scale Invariant Feature extraction (SIFT) is used for necessary feature extraction. Support Vector Machine is used as a classifier.

Lalit Verma et al. [3] implanted a hand gesture based system for human computer interaction. In system flow of this paper is hand gesture acquisition, segmentation, filtering, representation and classification is developed. Contour is employed as a feature extraction. Contour of the hand is represented by localized contour sequences. American Sign Language (ASL) is employed as a Database for training purpose. Coding is done in C+ +.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 09
September 2020

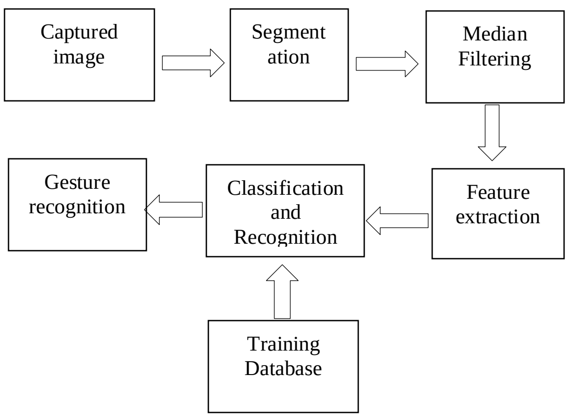
MA Gengyu et al. [5] propose a skin color segmentation for hand region segmentation. Distribution feature measurement is described in this paper for hand shape in the images. A 3-D Object browser is developed. Hidden Markov Model is used for classification purpose. The described system could analyze 2 frames per second.

Tom Chau et al. [7] try to give a different method which is basically hierarchical-feature based on pixel Ada Boosting (PBHFA), colour segmentation which is skin pixel dependent and a methodology of cluttered background subtraction. This feature meaningfully diminishes the training time by a minimum of thousand-fold compared to the standard Haar-like feature. During this paper ,authors has tried to reduce the tracking time of hand gesture and training time and training image of hand gesture.

Macro Maisto et al. [8] describe an accurate algorithm for find out the fingertips.At first contour of the segmented hand is found out. Convex hull of the contour and Convexity defect of the contour is found out. The intersection of the convex hull and convexity defects gives the fingertips the hand. Image of the hand is captured using depth sensor camera for e.g. Kinect camera.

III. System Architecture

The system includes a Logitech c270 HD camera with 3 Megapixels, a personal computer with MATLAB 2015a installed. The format of the image data collected by camera is RGB image. After a series of operations, including transformation of RGB image into YCbCr format, noise reduction, positioning, segmentation and recognition, the gesture recognition results are obtained. Hand gesture recognition for sign gesture recognition is show below in Fig 1;



IV. System Flowchart

Fig. 3.2 shows the hand recognition system work flow. The description of the of the block diagram and system workflow are following-

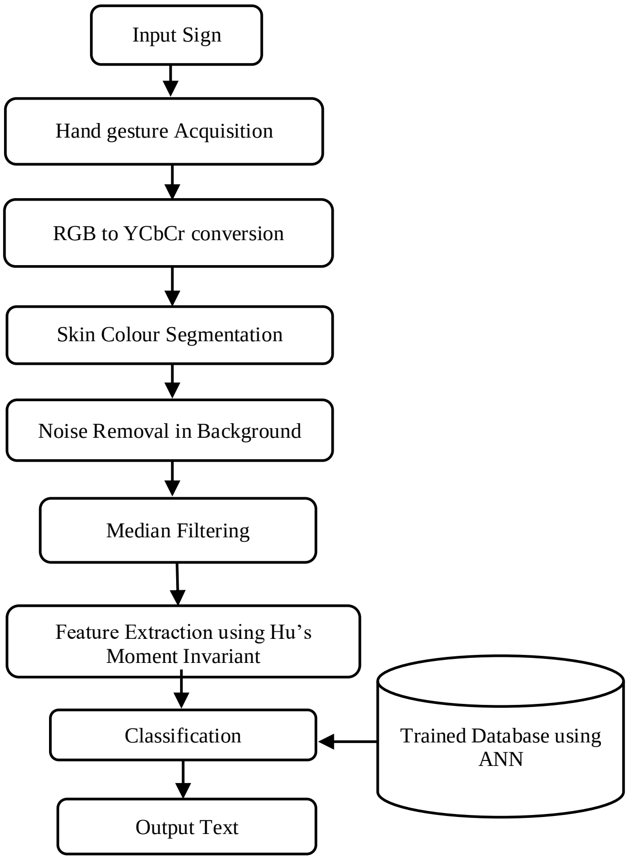


Fig.2. System flowchart of the hand gesture recognition

A. *Input Image capturing*

Color could be a robust feature. Image capturing is finished by different color space methods. It's susceptible to changing lighting conditions and it differs among people. During this proposed method, first of all, the snapshot of RGB image is captured by the three Megapixel web camera. Image are often given as an input within the style of Indian linguistic communication.

Fig.3. Hand gesture images as an input of the webcam in RGB colour format

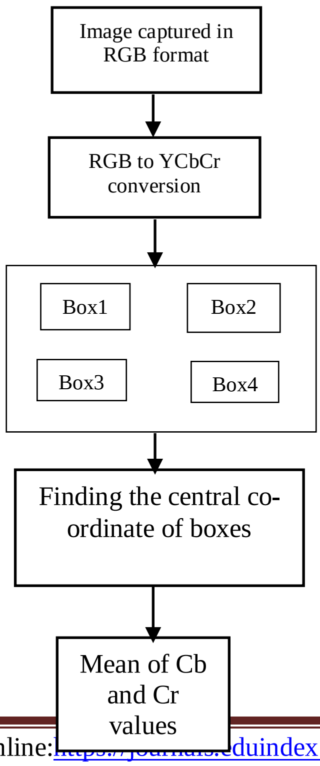
B. RGB to YCbCr conversion

Here the question arises that After all why we need to convert RGB into YCbCr image format?
The captured image is RGB format but medical research proved that the human eye has different sensitivity to colour and brightness. Thus there happen the transformation of RGB to YCbCr format.
RGB to YCbCr may be converted with the help of below given formula.

$$Y = 16 + (65.481*R + 128.553*G + 24.996*B)$$
$$Cb = 128 +(-37.997*R -74.203*G – 112.0*B)$$
$$Cr = 128 +(112*R – 93.786*G – 18.214*B)$$

Fig. 4. YCbCr images of "L" and "V"

C. Adaptive Skin colour segmentation



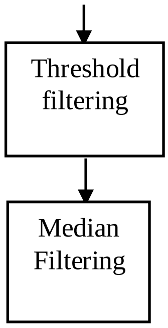


Fig.5. Adaptive skin colour segmentation

Fig.5. shows the system workflow of adaptive skin colour segmentation technique for human computer interaction. The description of system work flow are as following-

•The captured image through normal camera is generally in RGB format. In this format it is very difficult to separate the foreground image from background image.

•RGB image is converted into YCbCr format. Converting in the YCbCr format plane light compensation cab be achieved. The advantage of using this plane is luminance can be removed from the images. We have to set only the threshold value of the chrominance part.

•Creating the bounding box as per our needs. These bounding box edges can be darker with any colour. These bounding box indicates that we have to keep our fingers and palm on the boxes.

•After this finding the central coordinate of the bounding box because the region covered by the bounding box certainly would be skin colour part. In this proposed system all the central coordinate values of all six boxes is taken.

•Now finding the YCbCr values of central coordinates one by one we can find out the mean values of Cb and Cr. The reason behind for finding out the mean is we can set the threshold value.

•After finding out the mean of Cb and Cr it is more robust to set the threshold values by putting some deviations. One can put the deviations of 5%, so 0.95*Cb to 1.05*Cb and 0.95*Cr to 1.05*Cr is used to set the threshold value.

In the system, skin color model is used to separate "skin color area" from "no skin color area" which means that the overlap region of two areas is small as far as possible in the color space. By comparing the result of clustering in multiple color spaces, skin color in the RGB cluster is more compact and easier to implement clustering algorithm. Hence we use the YCbCr color space to detect skin color.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 09
September 2020

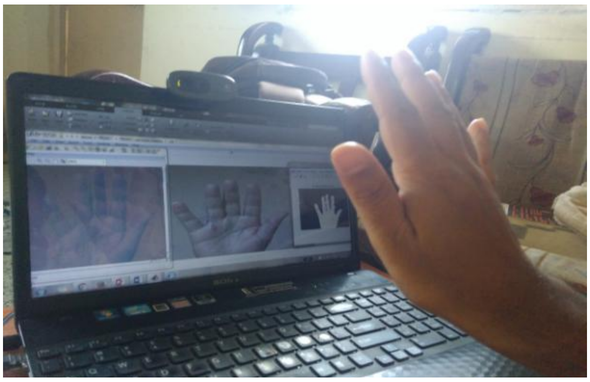


Fig.6. Experimental setup of hand separation from background using Adaptive skin colour segmentation

D. Median Filtering

After the resholding the YCbCr image, there are some noisy spot may be seen in the figure 2.Features values are affected with these noisy spot more than expected deviation. These spots also affects accuracy for the recognition of the system at the secondary subsequent stage. So these distortion is through 'Median' filtering. During this research 5x5 scalar (identity) matrix is employed to get rid of the noisy spot.

Original image Segmented Image Filtered Image

Original image Segmented Image Filtered Image

Fig. 7.Median filter Image output

The advantage of using Median Filter is 'skin-colour' region are often extracted from the captured image more accurately. So lastly, it can be said that colouring segmentation and median filter is stead fastly robust against in uniform background and lighting conditions.

E. Feature Extraction

Hu's Movement Invariant algorithm is employed as a feature extraction due to its effectiveness and usefulness in body identification. There is wide range of use of these algorithm in different types of classification likes airplanes, vessels, gesture and leaf recognition etc.

Hu's moment is defined seven functions, calculated from central moments of orders two, which were invariant with respect to translation, rotation and scaling of object. Hu'smoment invariants is calculated as below

$\Omega_1 = \eta_{20} + \eta_{02}$ (1)

$\Omega_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2$ (2)

$\Omega_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$ (3)

$\Omega_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$ (4)

$\Omega_5 = (\eta_{30} - 3\eta_{12})\,(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})$
$[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$ (5)

$\Omega_6 = (\eta_{20} + \eta_{02})\,[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})^2(\eta_{21} + \eta_{03})$ (6)

$\Omega_7 = (3\eta_{21} - \eta_{03})\,(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2$-$3(\eta_{21} +$ *η03)2*](*3η12−η30) (η21+* *η03)*[3*(η30+η12)2*-
$(\eta_{21} + \eta_{03})^2$(7)

Where η_{ab} is the stabilized central moment of order (a + b), which is defined as

$$\eta_{ab} = \frac{\varphi_{ab}}{\varphi_{00}{}^{\gamma}},$$

$$\Psi = \frac{a+b}{2} + 1$$

φ_{ab} is the parallel central moment, defined as

$$\varphi_{ab} = \sum_x \sum_y (p - \overline{q})^a (q - \overline{q})^b\, f(p,q),$$

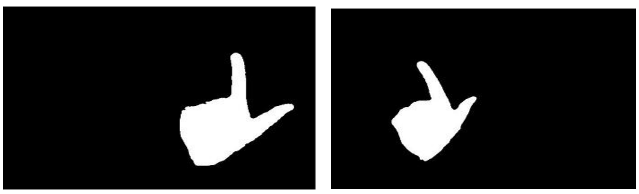
$$\overline{p} = \frac{m_{10}}{m_{00}} \text{ And } \overline{q} = \frac{m_{01}}{m_{00}}.$$

m_{ab} is two dimensional moment of order (a+b), well-defined as

$$m_{ab} = \sum_p \sum_q p^a q^b\, f(p,q).$$

Ω_1 - Ω_6 are moment invariants with respect to revolution and reflection, while Ω_7 fluctuates sign under reflection.

Example of Moment invariants



ISL of "L" Rotated and Scaled Image

International Journal of Research
Available at
https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 09
September 2020

ISL of "V" Rotated and Scaled Image

Fig. 8. Different segments of images hand gesture as Normal, Rotated and Scaled .

Same set of gesture have different sizes in terms of scaling, rotation and translation as shown in Fig. 8.

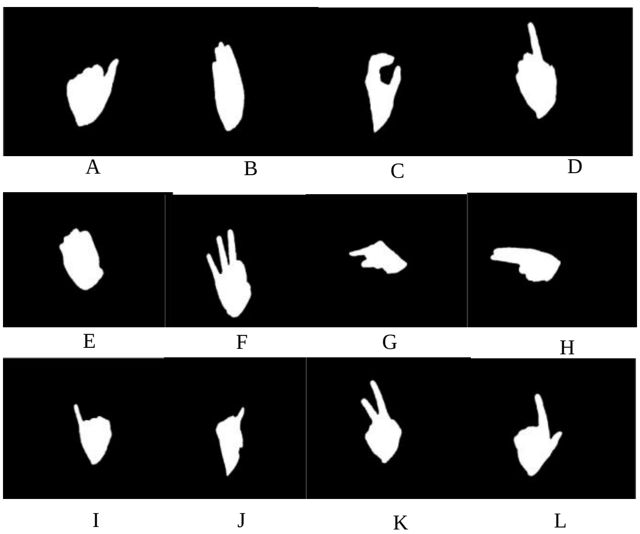
Table I. Moment Invariants values of Fig. 8

Ω	L	L_rot	V	V_rot
Ω_1	0.22729	0.21784	0.27156	0.26907
Ω_2	0.00705	0.00839	0.03693	0.03645
Ω_3	0.00106	0.00113	0.00225	0.00236
Ω_4	0.00114	0.00093	0.00186	0.00179
Ω_5	5.92E-07	5.54E-07	3.02E-06	5.23E-06
Ω_6	9.06E-05	7.78E-05	0.00035	0.00034
Ω_7	1.10E-06	3.79E-07	-2.07E-07	-2.23E-07

Table I shows that Hu's moment value is invariant to the Gestures rotation, scaling and translation. It is seen that first four moment value i.e. $\Omega_1, \Omega_2, \Omega_3$ and Ω_4 is being changed slightly approx. 2% variations from normal image while next three may be discarded for seeing. So as a result feature extraction using moment invariant are often progressed rapidly and efficiently.

F. Database for Training using (ANN

Database consists of a no. of images for every gesture. In this system each 26 alphabet gesture has been accepted. This database usually follow the Indian Sign Language.



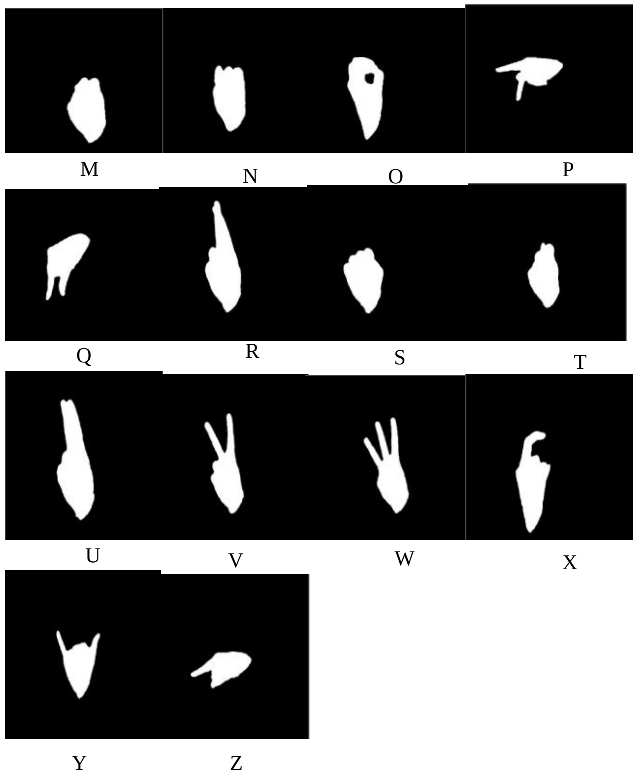


Fig. 9 Database Images of ISL for English alphabets

1. Training Images

It contains 84 images for each alphabet i.e. total number of images for training is 84x26=2184. All the images are trained using Artificial Neural Network (ANN) pattern recognition network with 600 hidden layer.

2. Testing Images

It contains 42 images for each alphabet for testing and again training so that accuracy can be improved.

3. Validation Images

It contains 84 images for each alphabet for first time for testing to validate the system.

G. Gesture Classification

After training of the system the only work is input given to the system. This input is nothing but the testing images. Testing images contains 26 set for each alphabet. Each set contains 42 images. These images are used for testing purpose.

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 09
September 2020

Taking “L” and “V” as input

Table II Output value after comparing with trained database

Input Image	Classification with database	Classification with database	Output for “L”	Output for “V”
A	1.87E-15	2.37E-22	0	0
B	5.25E-14	4.66E-24	0	0
C	3.60E-09	5.72E-25	0	0
D	0.000164	6.44E-16	0	0
E	2.83E-37	1.06E-25	0	0
F	3.65E-06	1.65E-16	0	0
G	1.03E-09	7.67E-32	0	0
H	4.80E-11	2.59E-12	0	0
I	9.26E-11	1.92E-21	0	0
J	2.79E-16	6.95E-25	0	0
K	0.000198	1.82E-13	0	0
L(input)	**0.999635 (max)**	8.92E-15	**1**	0
M	1.87E-14	3.17E-38	0	0
N	1.95E-16	1.46E-22	0	0
O	3.98E-15	9.13E-27	0	0
P	1.90E-09	5.74E-20	0	0
Q	6.28E-11	1.35E-22	0	0
R	2.93E-11	1.49E-14	0	0
S	1.06E-20	1.76E-34	0	0
T	6.31E-16	4.01E-32	0	0
U	1.77E-14	1.27E-13	0	0
V(input)	8.17E-11	**1.00000 (max)**	0	**1**
W	5.29E-12	1.52E-17	0	0
X	2.33E-09	5.66E-24	0	0
Y	8.77E-18	5.62E-32	0	0
Z	2.60E-12	8.35E-19	0	0

Images shown in Fig3.3 are given as input to the system after conversion of YCbCr images in fig. 3.4, filter images is obtained in fig. 3.7. In Table 3.1 Feature value is extracted through Hu’s Moment invariant method. Now these value is compared with trained database and following value is obtained through simulation dynamic link command in Matlab 2015a.

In Table III it can be seen that if “L” set of images is given to the system then value of at L is maximum and if “V” set of images is given to the system then it is compared with trained database and the output value is obtained. So it can be seen that value at V is maximum is obtained. So through following algorithm binary values can be obtained-

If (variable > 0.5)

Then 1 is obtained

Else

0 is obtained.

Let x and y be the variables. So

x = [0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

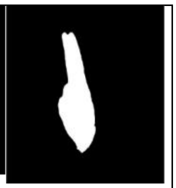
↑ L

y = [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]

↑ V

The value of x and y is compared with classification matrix and it gives the following output- x = L and y = V.

H. ACCURACY

Hand Gesture			
Classification accuracy	99%	99%	98%

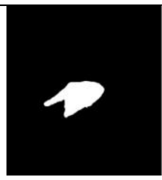
			
Classification accuracy	94%	94%	95%

Fig. 10. Classification Accuracy

The Fig.10. is showing the classification of gesture movement accuracy.

In the early days, artificial neural network (ANN) is basically explained with 2184 samples of images which were representing 26 hand images which means there is a hand image for each alphabet and each image is taken under stationary uniform background. Each gesture approximately has 84 different images.

Input is single feature vector which is given to Artificial Neural Network. For example, detecting ‘A’ feature vector is [1 0 0 0 0 0……0], for detecting ‘B’ feature vector is [0 1 0 0 0 …..0] and so on.

V. Conclusion and Future Scope

The main aim of this system is to recognize the hand gesture with more accuracy. The hand of different sizes and different in colour vary in races can be recognized through adaptive skin colour segmentation. Light variations is no more big issues if skin colour segmentation is used. All the alphabet gesture except “O” and “C” can be recognized with accuracy more than 90%. The system is robust against cluttered background. 2184 images of database is trained, each set carry 84 images. 42 images for testing purpose and 42 for validation purpose is kept. The future scope of this system is real-time implementation. So it is important to find out very robust feature for e.g. Fingertip detection.

REFERENCES

[1] Q. Nguyen and P. Premaratne, "Consumer electronics control system based on hand gesture moment invariants,"IET computer vision,, April 2007, pp. 35-41.

[2] Nicolas D. Georganas and Nasser H. Dardas,"Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques,"IEEE Transaction on Instumentation and Measurement, vol. 60, no. 11, nov 2011, pp 3592-3607.

[3] Suwei Ma and Lalit Gupta, "Gesture based interaction and Communication: Automated Classifier of Hand Gesture Contours,"Ieee trans. On system, mans and cybernatics, vol 31, no. 1, Feb 2001, pp. 114-120.

[4] T.Y. Kim and C. Nyirarugira, "Adaptive Differential Evolution Algorithm for Real Time Object Tracking," IEEE Transactions on Consumer Electronics, Vol. 59, No. 4, November 2013, pp. 833-838.

[5] Lin Xueyin and Ma Gengyu, "Operational Gesture Segmentation and Recognition,"Xsinghua science and technology,"Volume 8, Number 2, April 2003, pp.l69 – 173.

[6] Yun-Fu Liu and Jing-Ming Guo, "Improved Hand Tracking System,"IEEE Transaction on circuits and systems for video technology,"vol. 22, no. 5, May 2012, pp.693-701.

[7] Tom Chau and Natasha Alves, "Vision-Based Segmentation of Continuous Mechanomyographic Grasping Sequences," IEEE Transaction on biomedical engineering, vol. 55, no. 2, Feb. 2008, pp. 765-773.

[8] Marco Maisto, Massimo Panella, Luca Liparulo, and Andrea Proietti, "An Accurate Algorithm for the Identification of Fingertips Using an RGB-D Camera" IEEE journal on emerging and selected topics in circuits and systems, vol. 3, no. 2, June 2013

[9] P. Viola and M. Jones, "Robust real-time object detection," Int. Jour. Compute. Vis., vol. 2, no. 57, pp. 137–154, 2004.

[10] D. G. Lowe, "Distinctive image features from scale-invariant key points,"Int. J. Compute. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[11] D. G. Lowe, "Distinctive image features from scale-invariant key points,"Int. J. Compute. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[12] A. El-Sawah, N. Georganas, and E. Petriu, "A prototype for 3-D hand tracking and gesture estimation," IEEE Trans. Instrum. Meas., vol. 57, no. 8, pp. 1627–1636, Aug. 2008.