ISSN: 2393-8528

Contents lists available at www.ijicse.in

International Journal of Innovative Computer Science & Engineering

Volume 4 Issue 2; March-April-2017; Page No. 33-41

Static Hand Gesture Recognition using Mon-vision Based Techniques

N.S Sreekanth[1], N.K Narayanan[2],

[1]C-DAC Bangalore, #68 Electronics City, Bangalore 560100, India

nss@cdac.in

[2]Department of Information Technology, Kannur University, Kannur, Kerala, India

nknarayanan@gmail.com

Received 07 March. 2017; Accepted 16 April. 2017

ABSTRACT

This paper reports the experimental studies conducted for recognizing the hand based static gestures. Mono-vision based skin color segmentation techniques are used for segmenting the hand form a complex image sequence. The standard histogram features along with various geometrical features are extracted from the experimental context under uniform lighting conditions. American sign Language based hand gestures corresponding to the digits, ten digits, i.e., zero-to nine are used for study. Radial basis function neural network are used for classification.

Key Words: gesture recognition, static gesture recognition, mono-vision

INTRODUCTION

In human communication scenario most of the time it is found that, gestures are co-expressive with speech. Sometimes it also exists as an independent mode for human communication especially in the noisy environment or due to the disabilities. Automatic recognition of gesture by machine is being a topic of interest for researchers for quite a long time. The gesture recognition opened up new mode of interaction in Human Computer Interface (HCI). Gestures have long been considered as an interaction technique that can potentially deliver more natural, creative and intuitive methods for communicating with computers [1]. The gestures are commonly used in human-human interaction and play a major role in communication when the participants are unable to speak, or the situation does not allow the participant to speak etc. The gestures also play as an offset-input to the other mode of communication, for example gesture and speech are co-expressive and they form a part of rich human conversational features [2]. Gestures allow individuals to communicate a variety of feelings and thoughts, dislike, approval and affection, often together with body language in addition to words when they speak. In certain cases, we also use some object to show gestures; for example with a stick, the one who leads the orchestra he shows the gestures with help of stick to ensure the coordination with in the team. Gestures are broadly classified in to two; they are static gestures and dynamic gestures. **Static gestures** communicated via some symbols or fixed orientation of human body / body parts (hand, head) in particular manner. Generally sign languages are the subset of the static gestures. In static gestures generally no movement of body part is involved rather the orientation of human body or body part is involved in the gesture production. For example, to show gesture corresponding to "ok" we can keep our thumps up position by folding other fingers. **Dynamic gestures** as its name implies it involves a moving body part, i.e., hand, head, leg etc. For example, to give instruction for moving up we may use hand, which moves from bottom to top. In present work we address the recognition of static gesture recognition.

II. REVIEW OF PREVIOUS WORK

Automatic gesture recognition is addressed in various ways; starting from hardware based solution to machine vision based soft solution. If we look at the evolution of gesture based interaction, initially there were glove based devices, but they lacked the naturalness factor as they had introduced an additional hardware constraint on the user. Data Glove based approach uses a glove-type device which is mounted with various sensors to capture

**Corresponding author: N.S Sreekanth |*

movement of fingers and hand; and also their orientations in the space. These approaches can easily provide exact coordinates of palm and finger's location, orientation, and hand configurations [3] , [4], [5] .The main advantage of these approaches are high accuracy and fast reaction speed but this approach can be quite expensive. The modern techniques employed for gesture processing are either 3D models or image based processing [6]. The former lacks the computational efficiency and the simplicity compared to other. In the image based processing method, there are several techniques based on color, contour and correlation for identifying gestures [3], [7]. The work done by Gupta, et.al., reports a colour and contour based classification of hand gesture which can be used as alternative and augmentative communication for human machine interaction [8]. The experimental studies carried out by Mohandes, et al., reports a machine vision based approach combined with sensor based technique for recognizing Arabic Sign Language [9]. Another interesting study reported by Ueda., et al., extract the joint angles of hand and hand regions from multiple images captured via multi-view point cameras. In this method they integrate these multi viewpoint silhouette images, a hand pose is reconstructed as a "voxel model." Then, all joint angles are estimated using a three-dimensional model fitting between the hand model and the voxel model [10]. As this method uses multiple cameras to capture the hand gestures, it ensure more reliable results but the implementation will become difficult both in terms of computational complexities and cost of affordability for practical applications. In this view lots of popularities are gained for implementing the gesture recognition system with mono-vision techniques. One of the interesting practical application built by Habib and Mufti, report the details about the implementation of a mono-vision based virtual key board designed for interacting with computers, PDA's and mobile phones [11].

In this paper we have focused on the mono-vision (single camera) based colour-contour based techniques to recognize the hand gestures. The predefined orientation of fingers / hand in space is captured and relevant feature vectors are extracted for recognizing the gestures. The geometrical features are extracted from the captured gestures.

III. STATIC GESTURE RECOGNITION - MONO-VISION BASED APPROACH

There are various machine vision based techniques to process and recognize the gestures. Mono-vision based technique is more popular among these methods because of the less-computational complexities and cost-effective implementation for practical use. In mono vision based technique the images or video from a single camera is considered for processing and extraction of relevant gestures. In this paper mono-vision colour-contour based static gesture recognition is discussed. The user has to make gesture in front of a camera and the system captures the digital image. The captured image will be processed for segmenting the hand from the complex background. The segmentation of hand from the complex image is being done using the colour band based segmentation technique. Skin color segmentation algorithm identifies the skin colour region from the acquired image and separates the background from the image other than skin color regions. After the segmentation operation the relevant geometrical features are extracted from the segmented image for recognition. Appropriate classification algorithms can be used for classifying the gestures. The overall work flow for building a static gesture recognition system is shown Fig 1.

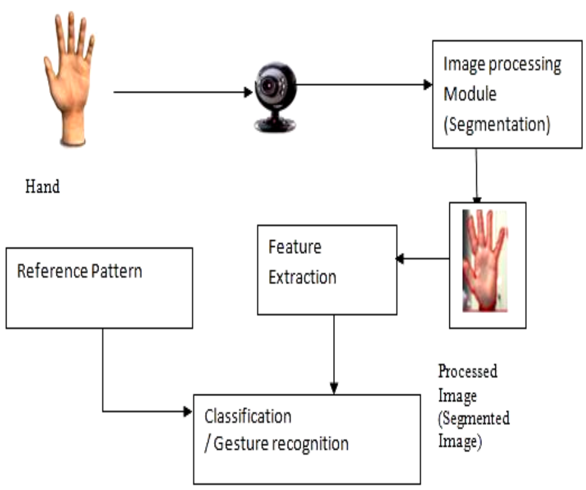


Figure1: Over all work flow for Gesture Recognition system

IV.SEGMENTATION - SKIN COLOUR BASED TECHNIQUES.

Skin color is one of most important source of information widely used in Human-Computer Interaction (HCI). The skin colour based human body part segmentation from a complex image background is extensively studied and reliable results are reported across the literatures [12]. The work done by Lee and Park, designed a vision based remote control system where colour band (skin color based) segmentation techniques are used to segment the hand and fingers from the complex image background. The gestures produced by hand will be converted in to commands to control electronic devices [13]. Ma and Lin developed a method using a skin-color-based technique to extract the area representing the hand in a single image with a distribution feature measurement designed to describe the hand shape in the images [14]. The studies on automatic sign language recognition reported by Han J, et al., used skin colour based segmentation technique where the skin colour model is built based on the combination of support vector machine active learning and region segmentation [15]. There are many colour space models for skin colour detection. The various television standards, computer graphics and video transmission gave birth for various colour space models. The most popular models are RGB model, Normalized RGB, HSI, YCrCb. Among these colour space models, RGB based methods gained lot of popularity because of computational simplicity over other methods and it is the widely used color spaces for storing of digital image data. [16], [17], [18]. In this paper we also used the RGB based colour space models for static gesture recognition.

Pixel based skin colour segmentation are simulated as part of this study where each pixel will be classified as skin or non skin pixel based on the color band identified against skin colour. Individual pixel classification may give many false positive regions. In order to avoid such things there has been lots of studies around the region based segmentation, where geometrical properties of object to be segmented will be considered. Region-based methods try to take the spatial arrangement of skin pixels into account during the detection stage to enhance the performance [19], [20], [21]. In RGB colour space model, we have to come up with the range of RGB values where human skin falls. As discussed earlier each pixel has to be identified as either skin pixel or non-skin pixel. There are many skin color types, ranging from whitish, yellowish, blackish and brownish, which must be all classified in one class, i.e., skin color. The decision rule for classifying a given pixel as skin or non-skin is being done using the following decision rule equitation. 6.1. [22], [23].

$$
\begin{aligned}
&(R,G,B) \text{ is classified as skin if:}\\
&R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and}\\
&\max\{R,G,B\} - \min\{R,G,B\} > 15 \text{ and}\\
&|R - G| > 15 \text{ and } R > G \text{ and } R > B
\end{aligned}
$$

The simplicity of this method attracted many researchers as it lead to a computationally simple and rapid classifier. In this work we have used the RGB colour space for skin modeling with above said decision rules. The pixels in the image frames are analyzed and the pixels which fall in to the above said conditions are mapped to skin colour else to non-skin colour and convert it as black pixel. The elimination of false positive can be done by checking the adjacent eight pixels with respect to current pixels. If at least four pixels are not within the above said range it will be treated as noise and will be eliminated. The images of ASL based gesture, digit '5' segmented using RGB based skin modeling are given in Fig.2.

Figure 2: ASL gesture digit "5" with hand and it's segmented Image using RGB Colour Segmentation

V. FEATURE EXTRACTION FOR GESTURE RECOGNITION

After the execution of segmentation algorithm, skin colour pixels will be highlighted and non-skin pixel regions can be neglected or can be converted as black. Now we have to extract the relevant features from the segmented image for performing the recognition operation. In order to identify the orientation of hand and fingers we use various geometrical properties of segmented image, i.e., hand. First, the smallest bounding rectangle is constructed around the segmented hand, then apply popular convex hull algorithm to construct a smallest polygon which covers all parts of the segmented hand. The polygon return by the convex hull algorithm will be a subset of bounding rectangle. The convex hull algorithm is shown below [32].

Algorithm ConvexHull(P)

Input. A set P of points in the plane.

Output. A list containing the vertices of CH(P) in clockwise order.

1. Sort the points by x-coordinate, resulting in a sequence $p_1,\ldots,p_n$.
2. Put the points p_1 and p_2 in a linked list $\mathcal{L}_{upper}$, with p_1 as the first point.
3. **for** $i \leftarrow 3$ to n
4. **do** Append p_i to $\mathcal{L}_{upper}$,.
5. **while** $\mathcal{L}_{upper}$, contains more than two points and the last three points in $\mathcal{L}_{upper}$ do not make a right turn
6. **do** Delete the middle of the last three points from $\mathcal{L}_{upper}$.
7. Put the points p_n and p_{n-1} in a linked list $\mathcal{L}_{lower}$ with p_n as the first point.
8. **for** $i \leftarrow n-2$ down to 1
9. **do** Append p_i to $\mathcal{L}_{lower}$
10. **while** $\mathcal{L}_{lower}$ contains more than 2 points **and** the last three points in $\mathcal{L}_{lower}$ do not make a right turn
11. **do** Delete the middle of the last three points from $\mathcal{L}_{lower}$
12. Remove the first and last point from $\mathcal{L}_{lower}$ to avoid duplication of the points where the upper and lower hull meet.
13. Append $\mathcal{L}_{lower}$ to $\mathcal{L}_{upper}$ and call the resulting list L.
14. return L

The resultant polygon for a typical example is shown in Fig 3.

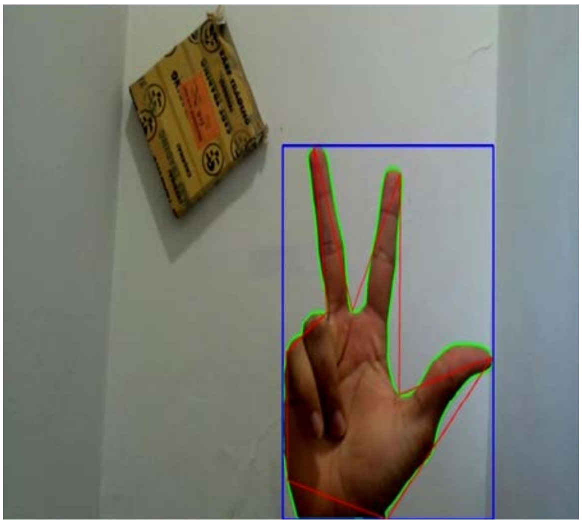


Figure3: After applying the Convex Hull algorithm for ASL based Gesture digit 3.

These features along with various other geometrical properties extracted from the processed image are used for gesture recognition. The set of geometrical features which are derived from the segmented and processed image are listed in Fig. 4. The images in RGB colour space can be mapped to gray scale image for further processing. The histogram coefficient of gray scale images has already proved as a reliable feature for recognizing static gestures [24], [25], [26] [27]. The geometrical features along with histogram coefficient are used for static gesture recognition are listed in below.

1. area -- gives the area of the region
2. perimeter -- gives the perimeter of the region
3. centroid -- gives the centroid of the region as a tuple (x,y)
4. bounding_box -- gives the bounding box parameters as a tuple => (x,y,width,height)
5. aspect_ratio -- aspect ratio is the ratio of width to height
6. equi_diameter -equivalent diameter of the circle with same as area as that of region
7. extent -- extent = contour area/bounding box area
8. convex_hull -- gives the convex hull of the region
9. convex_area -- gives the area of the convex hull
10. solidity -- solidity = contour area / convex hull area
11. center -- gives the center of the ellipse
12. majoraxis_length -- gives the length of major axis
13. minoraxis_length -- gives the length of minor axis
14. orientation -- gives the orientation of ellipse (inclination from horizontal axis)
15. filledImage -- returns the image where region is white and others are black
16. filledArea -- finds the number of white pixels in filled Image
17. convexImage -- returns the image where convex hull region is white and others are black
18. pixelList -- array of indices of on-pixels in filled Image

Figure 4: List of geometrical features

Fig 5. shows few examples captured from the experimental context, i.e, the bounding rectangle and smallest polygon obtained using convex hull algorithm. The resulted features extracted from the segmented image are also shown on the right of the image.

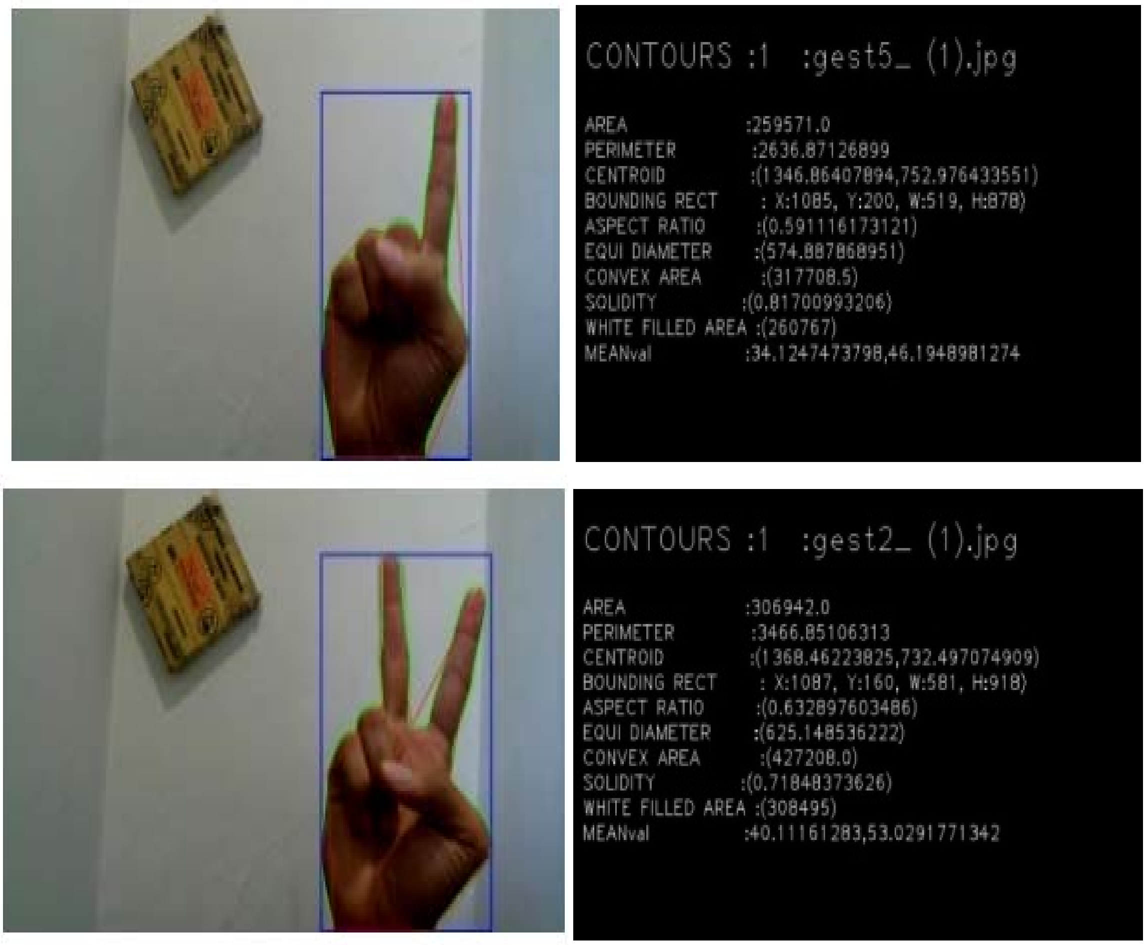


Figure 5: Segmentation of hand with Bounding rectangle and polygon.

Various geometrical features extracted are shown on the right side of the image.

VI. STATIC GESTURE RECOGNITION

Various features extracted from the segmented image will be given to classification algorithms to recognize/ classify the gestures. We have implemented k-mean based radial basis function neural network, for classifying the static gestures. The k-mean based Radial Basis Function Neural Network (RBFNN) is reported as an efficient method for various classification problems especially in gesture recognition and image classification applications [28] , [29], [30],[31]. Radial basis function neural network (RBFNN) is the most popular neural network (NN) because it has a simple architecture, only one hidden layer between input and output layer. In hidden layer, localized radial basis function are used to transform the feature vector from input space to hidden space. Sensitivity of the hidden neuron is tuned by adjusting spread factor of basis function. This network is faster and free from local minima problem. In this work, the accurate centers of the RBFNN are chosen using k-means clustering algorithm [31]. The k-means RBFNN classification procedure is shown in below.

k-mean RBFNN Algorithm

- Step 1: Set value of m = total number of training classes and $i = 1$.
- Step 2: Take D_i ,the data of i_{th} class.
- Step 3: Set value of k = number of clusters in D_i .
- Step 4: Assign initial centroid of each cluster by selecting randomly k number of samples from data D_i or first k number of samples from data D_i .
- Step 5: Take each sample of data D_i and compute its Euclidean distant from centroid of each of the cluster and assign to the cluster with the nearest centroid.
- Step 6: Compute the centroid of each cluster.
- Step 7: Repeat step 5 to step 6 until centroid of the each cluster don't change.
- Step 8: Store k number of centroid as centers of RBFNN in i_{th} iteration and increase i by 1.
- Step 9: Repeat step 2 to step 8 until i <= m.

VII. SIMULATION EXPERIMENT AND RESULTS

The recognition/ classification experiments with static gestures are conducted using the k-mean Radial Basis Function Neural Network. Digits shown with hand, ASL based representation of digits (0 to 9) are experimented. Snapshot of data base is shown in Fig.5.Training data base is created with 28 individuals (15 male and 13 female) and for each class 30 sample images are collected from each user. Both training testing are carried out in uniform lightening conditions. The recognition accuracy of different gestures with 25 test samples for each class corresponding to digits is shown in Table 1.

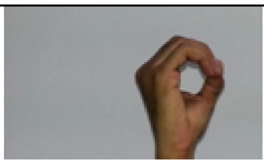
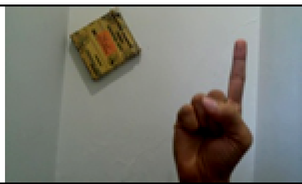
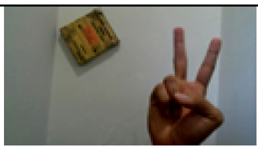
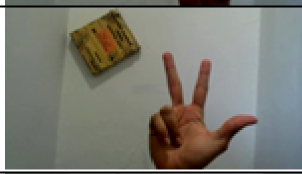
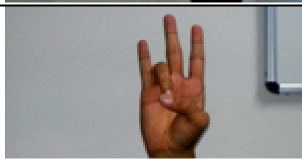
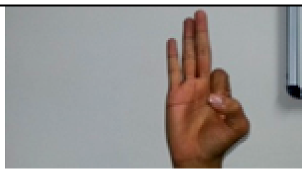
Digits	Gestures	Digits	Gestures
Zero		One	
Two		Three	
Four		Five	
Six		Seven	
Eight		Nine	

Figure 5: Gesture Data Base (Digits) considered for the experiments

Table 1: Recognition accuracy of Static gestures corresponding to Digits

Digits	Recognition Accuracy in %	Digits	Recognition Accuracy in %
One	98	Six	92
Two	95	Seven	89
Three	96	Eight	94
Four	97	Nine	87
Five	98	zero	89

VII. CONCLUSIONS

The ASL based static hand gestures are recognized using the mono-vision based techniques. Skin colour based segmentation techniques are used for segmenting the hand from the complex. The experiment is conducted under uniform lighting conditions. Along with standard histogram feature, geometrical features are extracted from the segmented image. ASL digits (zero to nine) are considered for the experiment. The studies report a fair result under said experimental conditions. Presence of skin colour other than the hand as part of the image will reduce the recognition accuracy. This can be addressed as a future study. The studies can also be conducted under different lighting conditions.

REFERENCES:

1. Siddharth S. Rautaray · Anupam Agrawal "Vision based hand gesture recognition for human computer interaction: a survey", Artificial Intelligence Review , Springer , Volume 43, Issue 1 , pp 1-54., 2012
2. Francis Quek, David McNeilly, Robert Bryll, Susan Duncan, Xin-Feng Ma, Cemil Kirbas, Karl E. McCulloughy, and Rashid Ansari : Gesture and Speech multimodal conversational interaction. ACM Transactions on Computer-Human Interaction, Vol. 9, No. 3, September 2002, Pages 171–193, 2002.
3. Harshith.C, Karthik.R.Shastry, Manoj Ravindran, M.V.V.N.S Srikanth, Navee Lakshmikhanth "Survey on various Gesture Recognition Techniques for Interfacing Machines based on Ambient Intelligence" (IJCSES) Vol.1, No.2, November 2010.
4. PragatiGarg, Naveen Aggarwal and SanjeevSofat, "Vision Based Hand Gesture Recognition, World Academy of Science, Engineering and Technology 49, pp. 972-977., 2009
5. Laura Dipietro, Angelo M. Sabatini, and Paolo Dario, Survey of Glove-Based Systems and their applications, IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 38(4), 2008 pp. 461-482, doi: 10.1109/TSMCC.2008.923862., 2008.
6. Pavlovic, V.I. Sharma, R. ; Huang, T.S. "Visual interpretation of hand gestures for human-computer interaction: a review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume:19 , Issue: 7 , pp. 677-695., , 2002.
7. Mutra. S, Acharya T., "Gesture Recognition - A survey", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Volume:37 , Issue: 3, pp. 311 - 324, 2007 .
8. Gupta, L. ; Suwei Ma, "Gesture-based interaction and communication: automated classification of hand gesture contours", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Volume:31 , Issue: 1,2001, pp.114-120, 2001.
9. Mohandes, M; Deriche, M. ; Liu, J., "Image-Based and Sensor-Based Approaches to Arabic Sign Language Recognition", IEEE Transactions on Human-Machine Systems, Volume:44 , Issue: 4 , pp. 551-557, 2014.
10. Ueda, E. Matsumoto, Y. ; Imai, M. ; Ogasawara, T. "A hand-pose estimation for vision-based human interfaces", IEEE Transactions on Industrial Electronics, Volume:50 , Issue: 4, pp.676-684, 2003..
11. Habib, H.A, Mufti, M. "Real time mono vision gesture based virtual keyboard system", IEEE Transactions on Consumer Electronics, Volume:52 , Issue: 4, , pp.1261-1266., 2007.
12. Nam Vo, Quang Tran ; Thang Ba Dinh ; Tien Ba Dinh ; Nguyen, Q.M." An Efficient Human-Computer Interaction Framework Using Skin Color Tracking and Gesture Recognition", Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), IEEE RIVF International Conference on , 1-4 Nov. 2010, Hanoi, pp.1-6., 2010

13. Daeho Lee, Youngtae Park, "Vision-based remote control system by motion detection and open finger counting", IEEE Transactions on Consumer Electronics,Volume:55, Issue: 4, pp. 2308 - 2313, 2010.

14. Ma, Gengyu, Lin, Xueyin " Operational gesture segmentation and recognition ", Tsinghua Science and Technology (Volume:8 , Issue: 2) 2012 Tsinghua University Press (TUP), pp. 169 – 173., 2012.

15. Han, J. Awad, G. ; Sutherland, A. "Automatic skin segmentation and tracking in sign language recognition" IET Computer Vision, Volume:3 , Issue: 1, pp.24-35., 2009.

16. Harpreet Kaur Saini, Onkar Chand "Skin Segmentation Using RGB Color Model and Implementation of Switching Conditions" International Journal of Engineering Research and Applications (IJERA) Vol. 3, Issue 1, January -February 2013, pp.1781-1787., 2013

17. Mokhtar M. Hasan, Pramod K. Mishra,"Novel algorithm for skin color based segmentation using mixture of GMMs" Signal & Image Processing : An International Journal (SIPIJ) Vol.4, No.4, August 2013, pp. 139-148., 2013

18. Omanovic, S. Buza, E; Besic, I, "RGB ratios based skin detection" , 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014, 26-30 May 2014, pp.1348 – 1353.,2014

19. Yang, M.-H., Ahuja, N. 1998. Detecting human faces in color images. In International Conference on Image Processing (ICIP), vol. 1, 127–130

20. Kruppa, H., Bauer, M. A.,Schiele, B. 2002. Skin patch detection in real-world images. In Annual Symposium for Pattern Recognition of the DAGM, Springer LNCS 2449, 109–117., 2002.

21. Jedynak, B., Zheng, H., Daoudi, M., And Barret, D. 2002. Maximum entropy models for skin detection. Tech. Rep. XIII, Universite desSciences et Technologies de Lille, France., 2002.

22. Fan Hai Xiang and Shahrel Azmin Suandi, "Fusion of Multi Color Space for Human Skin Region Segmentation", International Journal of Information and Electronics Engineering, Vol. 3, No. 2, March 2013, pp.172-174., 2013.

23. P. Peer and F. Solina, "An automatic human face detection method," in Proc. of 4th Computer Vision Winter Workshop (CVWW) .pp. 122-130, 1999.

24. Hanning Zhou; Lin, D.J. ; Huang, T.S., "Static Hand Gesture Recognition based on Local Orientation Histogram Feature Distribution Model", Conference on Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. 27-02 June 2004, pp.161-166., 2004.

25. Dardas, N.H, Georganas, Nicolas D., "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques " IEEE Transactions on Instrumentation and Measurement, Volume:60 , Issue: 11, 2011, pp. 3592 – 3607., 2011

26. Mihalache, C.R. Apostol, B., "Hand pose estimation using HOG features from RGB-D data ", 17th International Conference on System Theory, Control and Computing (ICSTCC), 11-13 Oct, 2013. 2013, pp. 356 – 361., 2013.

27. Viswanathan, D.M, Idicula, S.M., "Recognition of hand gestures of English alphabets using HOG method", Proc. International Conference on Data Science & Engineering (ICDSE), 2014., 26-28 Aug. 2014, pp. 219 – 223, 2014.

28. Camps-Valls, G.; Bruzzone, L."Kernel-based methods for hyperspectral image classification", IEEE Transactions on Geoscience and Remote Sensing, Volume: 43, Issue: 6, pp. 1351-1362, 2005

29. Christoyianni, I.; Dermatas, E.; Kokkinakis, G., "Fast detection of masses in computer-aided mammography", IEEE Signal Processing Magazine, Volume: 17, Issue: 1, pp.54-64, 2000

30. Jamuna Kanta Sing, , Dipak Kumar Basu, Mita Nasipuri, Mahantapas Kundu, "Face recognition using point symmetry distance-based RBF network" , Applied Soft Computing, Elsevier Publications, Volume 7, Issue 1, January 2007, Pages 58–70., 2007

31. Ghosh D.K, Ari S, "A static hand gesture recognition algorithm using k-mean based radial basis function neural network", 8th International Conference on Information, Communications and Signal Processing (ICICS) 2011, pp. 1-5.,2011.

32. Convex HULL pseudo code http://www.cse.iitm.ac.in/~augustine/cs6100_even2012/slides/03_ConvexHull.pdf