ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

Vision Based Human Computer Interaction Using Fingertip Tracking

Subhash Nevhal

Department of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India

ABSTRACT: Human Computer Interaction (HCI) is an active focus of research. Considering the motivation, need and challenges, here proposeda fingertip detection system with vision based approach for desktop handling operation. HCI exists everywhere in our daily lives. It is usually achieved by using a physical controller such as a mouse, keyboard or touch screen. It hinders Natural User Interface (NUI) as there is a big barrier between the user and computer. The design and development of a robust marker-less hand/finger tracking and gesture recognition system using low-cost hardware is presented here. A simple but efficient method that allows robust and fast hand tracking despite complex background and motion blur is devised. The system can translate the detected hands or gestures into different functional inputs and interfaces with other applications. It enables intuitive HCI and interactive motion gaming. The system is useful to avoid physical devices such as a keyboard, mouse and it can be used for remote-less control of the computer system, smart TV and interactive video-gaming.

KEYWORDS: Computer Vision, Human Computer Interaction, Fingertip detection, Hand tracking and Segmentation, Hand Gesture Recognition.

I. INTRODUCTION

Computer technologies have grown tremendously over some decades. As technologies are progressing rapidly, existing HCI techniques are becoming a bottleneck. The most common mode of HCI is relying on simple mechanical devices, i.e. keyboard and mouse. These devices are more familiar but are less natural and intuitive in interacting with computers. There has been much active research towards novel devices and techniques that allow gesture enabled HCI in recent years. There are generally two approaches to interpreting gestures in HCI by computers. First attempt to solve this problem resulted in a hardware-based approach. This approach requires a user to wear bulky devices, hindering ease and naturalness of interacting with the computer. Although the hardware-based approach provides high accuracy, it is not so practical. This has led to active research on more natural HCI technique, which is computer vision-based [13]. This approach uses cameras and computer vision techniques to interpret gestures. Research on vision-based HCI has enabled many new possibilities and interesting applications. Some of the most popular examples are tabletop, visual touchpad, and TV remote control. Vision-based HCI can be further categorized into the marker-based and marker-less approach. Several studies utilize color markers or gloves for real time hand tracking and gesture recognition. This approach is easier to implement and has better accuracy, but it is less natural and not intuitive. Other studies [15] focused on marker-less approach by using different techniques such as Haar-like features, Convexity defects, K-Curvature, Bag-of-features, Template Matching, Circular Hough Transform, Particle Filtering, and Hidden-Markov Model. Most studies on marker-less approach focused on recognize static hand poses, or dynamic gestures only, but not both. Several researchers used Haar-like features [15], which required high computing power. The classifier preparation stage also consumed a lot of time. Some studies used K-curvature to find peaks and valleys along a contour, and then classified these as fingertips. But, they did not solve the problem of differentiating between a human face and hand regions because they assumed that only hand regions are visible to the camera. Therefore, the problem when the hand is blocking in front of the face is also not discussed. Most of the studies only focused on efficient hand recognition algorithm but did not translate the detected hand into functional inputs. Some authors utilize static hand gesture recognition to simulate gaming inputs by translating different static gestures into keyboard events. While most of the researchers [16] develop sample application as a proof-of-concept, the hand tracking capability is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

limited to their application only and is not able to interface with other applications other than passing simple mouse events.

II. Literature Survey

There are many applications where fingertip tracking is required such as sign language, robotics, gaming, virtual reality etc. Ghotkar and Kharate [1], have implemented hand segmentation using Lab Color Space (HSL) and tested various techniques for efficient segmentation. They achieved better results on skin colour detection using HTS algorithm and used Mean shift algorithm as hand tracking technique. Ghotkar et.al. [2], presented gesture recognition system which took the input hand gestures through the in-built web camera. Hand tracking is done using the CAMShift method. The Segmentation of hands is carried out by using HSV colour model. The segmented hand image is represented using certain features. These features are used for gesture recognition using the Genetic Algorithm which gaveoptimized results. The result is converted into corresponding text and voice. HSV colour model is used as hand segmentation technique. General CAMShift algorithm is used for tracking. Wang et. al. [3], proposed a method, called the fingertip detection algorithm which is developed based on the prediction of the contents by using a projector and a binocular vision system based on two cameras. This is applied for detecting the depth of fingertips and touch operation. The prediction & triangulation method is used as hand segmentation technique. Akhalaq et. al. [4], proposed a hand gesture based human computer interaction system comprising of a web-camera and a pocket projector. The projector projects the display on the wall or any other plain surface. A user can interact with the projected screen using his fingertips which are tracked in the air by the camera using CAMShift tracker. Here skin colour distribution in the chrominance axis of the YCrCb colour space with the single Gaussian method is used as hand segmentation technique and CAMShift tracking algorithm is used as hand tracking technique. Chen et. al. [5], used a real-time hand tracking algorithm with a normal camera. KLT feature is used to tracking good features. This tracking result is utilized to calculate the main velocity of hand motion. Additionally, a global velocity which is calculated from a probability of Bayesian skin colour is used to refine the velocity of hand motion. Finally, CAMShift is used to detect a more precise hand region. Baraldi et. al. [6], proposed a method in which dense features are extracted from regions selected by a new hand segmentation technique. This can work in real time on a wearable device. Chien et. al. [7] proposed a method, which segmented hands out of an entire image and facilitated depth estimation of tracked hands in real-time by dual camera systems. Multi-threading and several techniques are applied to reduce computational complexity in proposed design. Dhote et. al. [8], used a method to segment hands out of an entire image and facilitated depth estimation of tracked hands in real-time by dual camera systems. Multi-threading and several techniques are applied to reduce computational complexity. Leonardo et. al. [12], presented a method for tracking and recognizing hand gestures by extracting unique invariant features from gestures. The extracted feature is used to perform effective matching between different observations of a hand gesture. They used RGB image based on the skin as hand segmentation technique and RGB with depth image as hand tracking technique. Nanivadekar et. al. [9], used YCbCr based skin colour model hand tracking technique. Here a three-step algorithm is used to get better quality hand tracking and segmentation. This algorithm worked on motion tracking, edge detection and skin colour detection.

III.Methodology

The proposed algorithm is useful to proper segmentation and efficient tracking of hand.

Firstly, image frames are acquired using camera feed. The hand is segmented using algorithm discussed below. The hand is tracked using CAMShift algorithm with Kalman filter. Hand centre is determined by using the largest contour of the segmented hand. These coordinates are used to set cursor position mapping. The virtual keyboard is used. Various hand gestures are mapped to mouse events such as click, double click etc.
The proposed algorithm is consisted of two main steps.

Step 1: Segmentation:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

- Blur the input frame to reduce noise.
- Convert frame to HSV.
- Separate out the part contained in the rectangle capture area and discard the rest.
- Calculate back projection of the frame.
- Apply morphology and smoothening techniques (Gaussian and Median blur) to the back projection generated.
- Apply threshold to generate a binary image from the back projection. This threshold is used as a mask to separate out the hand from the rest of the frame.

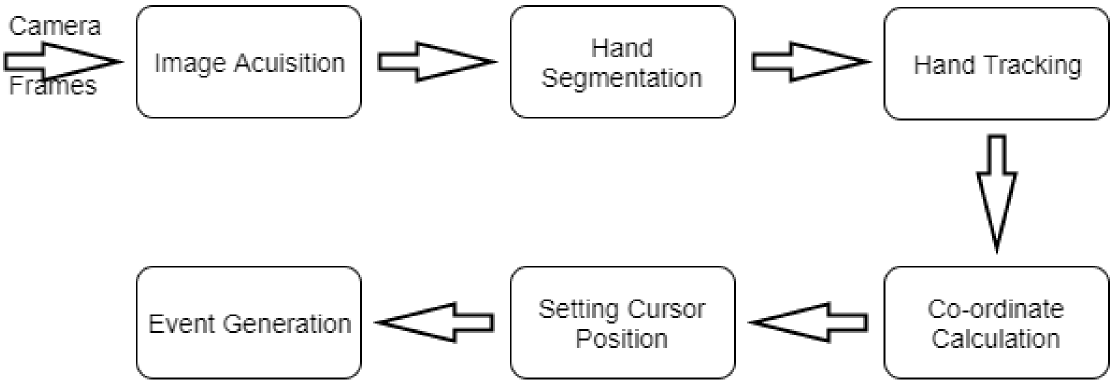

Fig. 1, System Architecture.

Step 2: Tracking:

The procedures of the CAMShift (Continuously Adaptive Mean Shift) algorithm [17,18] are as follows,

- According to the extracted target from the background subtraction, set the search window with a size of s in the color probability distribution.
- Calculate the zero-order matrix M_{00} of the search window

$$M_{00} = \sum_{i=x} \sum_{i=y} I(x,y)$$

The first-order matrix of x and y is

$$M_{10} = \sum_{i=x} \sum_{i=y} xI(x,y)$$

$$M_{01} = \sum_{i=x} \sum_{i=y} yI(x,y)$$

Where I(x,y) is the pixel value at the point (x,y), and the change range of (x,y) is inside the search windows.

- Calculate the centroid of the search window as the centroid of the target,

$$X_c = M_{10} / M_{00}$$

$$Y_c = M_{01} / M_{00}$$

Reset the search window size s as the color probability distribution function of the above search window.

- Repeat the above steps until convergence (centroid changes less than a given threshold).
- The major axis l, minor axis w and direction angle θ of the target can be obtained by calculating the second-order matrix.

$$M_{11} = \sum_{i=x} \sum_{i=y} xyI(x,y)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

$M_{20} = \sum_{i=x} \sum_{i=y} x^2 I(x,y)$

$M_{02} = \sum_{i=x} \sum_{i=y} y^2 I(x,y)$

$l = \frac{1}{2}\left[(a+c) + \sqrt{b^2 + (a-c)^2}\right]$

$w = \frac{1}{2}\left[(a+c) - \sqrt{b^2 + (a-c)^2}\right]$

$\theta = \frac{1}{2} arctan\left(\frac{2b}{a-c}\right)$

Where,

$a = \frac{M_{20}}{M_{00}} - x_c^2$; $b = 2\left(\frac{M_{11}}{M_{00}} - x_c y_c\right)$; $c = \frac{M_{02}}{M_{00}} - y_c^2$

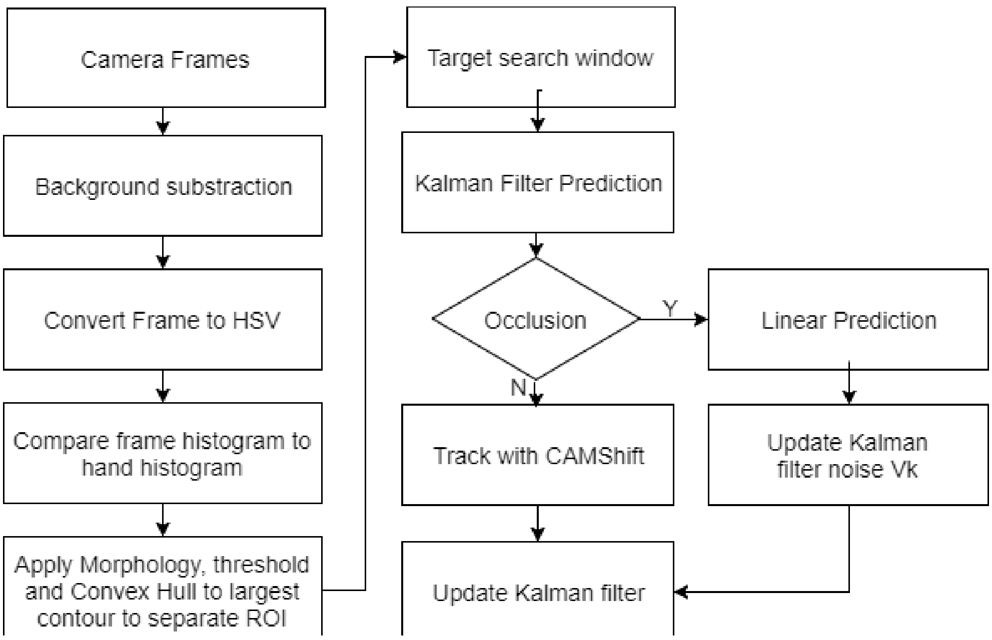

Fig 3, Flow chart of segmentation and tracking algorithm

Kalman Filter:

Kalman filter [18,19] is a linear recursive filter, it makes the optimal estimation of the next state based on the previous states sequence of the system and the prediction has characteristics as unbiased, stability and optimal. Kalman filter is a computational algorithm that implements a predictor-corrector type estimator to deduce optimum estimation of past, present and future state of a linear system in the sense that it minimizes the estimated error covariance. Therefore, by applying Kalman filter on the extracted hand locations, we can estimate and refine the optimal locations by removing unstable noises. It will result in smoother and less jumpy hand locations or trajectory paths that are more suitable for controlling applications. Kalman filter algorithm consists mainly of state and observation equations,

$X_c = A_k X_{k-1} + W_k$

$Z_k = H_k X_k + V_k$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

where X_k is the state vector, Z_k is the observed state vector of the system, A_k is the state transition matrix from time t_{k-1} to t_k, H_k is the observation matrix at time t_k, Dynamic noise Wk and the measurement noise V_k is uncorrelated white noise sequences.

The main steps of the CAMShift algorithm with Kalman filter are as follows

- Target location prediction in the current frame: The Kalman filter predicts the location of the target of the current frame. The Kalman filter prediction is based on objective historical movement information.
- Target matching: If the position of the moving target in the next image can be predicted, then the history of tracking positions and the positions unlikely to appear, can be filtered out in the next image. Depending upon color probability distribution of the target, use CAMShift algorithm to search the most similar goals with the target template in the neighbourhood of the Kalman filter prediction. Thus, the scope of the search can be reduced.
- Improvement: When meeting a large area of background interference which has the similar colour with the target, the target calculated by CAMShift algorithm expands rapidly, however, the normal deformation between the adjacent two frames varies slowly. If the difference between the adjacent two calculated target becomes larger than the threshold, at this time, the calculated target is dropped, instead of, the area predicted by Kalman filter is used to reinitialize the CAMShift algorithm. If the current target suddenly becomes much smaller than the last target out of the threshold, then it is considered that the target is obstructed. At this time, the prediction value calculated by Kalman filter is taken instead of the CAMShift algorithm and reinitialize it with the original detected target.
- Kalman filtering state Update: To update the state of the Kalman filter using the matched target location as the observed values of the Kalman filter, and thus estimate a more accurate predictive value of the next frame target.

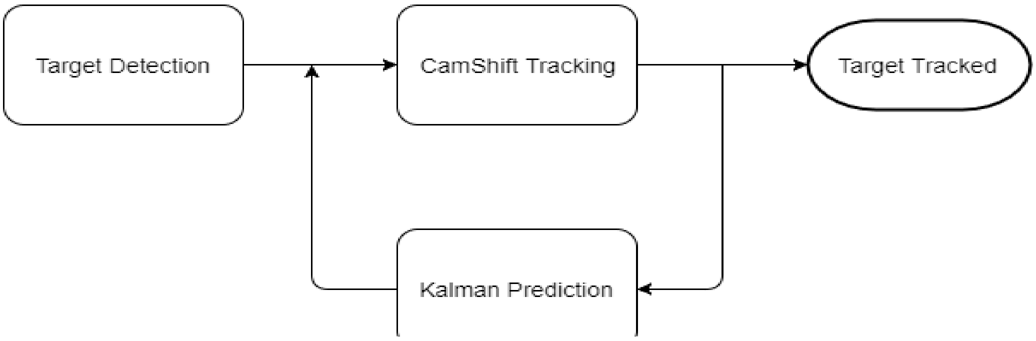

Fig.3, CAMShift tracker with Kalman filter.

Dataset: It comprised of various gestures which are stored in a dictionary. Thisdictionary is referred for the gesture recognition system. A gesture with name ‘V’is defined. The name is later used as an identifier or unique id to the gesture. Palmcentre is set as (475,225) with a radius of 45 pixels. This gesture is a hand showing2 fingers making a ‘V’. Another gesture with name ‘L right’ is defined. Palmcentre is set as (475,225) with a radius of 50 pixels. This gesture is a hand showing1 finger with a thumb making an inverted ‘L’. Next gesture is ‘Index Pointing’. HerePalm centre is set as (480,230) with a radius of 43 pixels. ‘Index Pointing’ is usedfor a mouse single click event.‘L right’ is used for a mouse double click event. ‘V’is used for a mouse right click event.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

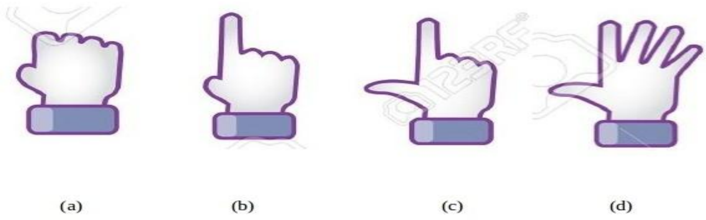

Fig 2, Gestures.

IV. MODULES AND DISCUSSIONS

The system can recognize different static gestures based on the proposed method. The skin color extraction method is similar to previous approaches, except for the usage of a narrower range. Despite using a narrow range, the face region is still extracted. Since it is based on feature detection, face removal will not work well if the hand is blocking the face region. This would cause a connected big contour to be extracted by our skin extraction method. Hence, background subtraction is applied to effectively separate the hand contour from the face contour. The hybrid hand and fingers tracking method is based on hand shape analysis that evolved from previous studies on Convexity Defects, K-Curvature and maximum inscribed circles. With this information, hand and fingertip locations, direction and thumb location can be found. One of the main advantages of the proposed solution is that users are not required to wear long sleeve shirts. Kalman filter is applied to stabilize the detected hand and fingers location. This makes the system more usable in applications that require precise control.

The system consists of three modules: A camera module, a detection module, and an interface module.

• Camera module: This module is responsible for connecting and capturing image output from different types of detectors (regular Web-cameras and Leap Motion cameras), and then processing this output with different image processing techniques. The output of this module is a smoothed binary image with clean contours suitable for hand detection. In this module, image frames are being retrieved from a USB Web-camera or Leap Motion Controller, andthen processed through several steps using image processing techniques. Then, the preprocessed image frame is passed to the next module for detection. The motivation was to design a system thatcan utilize low-cost USB Web-cameras, which are readily available in most user’s homes. However, as low-cost Leap Motion cameras are becoming more common nowadays, the support for Leap Motion cameras is also provided in the system. The Leap Motion camera is better in terms of hand segmentation. It is also significantly more expensive than a USB Web-camera.

• Detection module: This module is responsible for detection and tracking of hand and fingers. CAMShift algorithm and Kalman filter are applied to improve the accuracy and stability of tracking. The output of this module is hand and finger locations in 2D space. The output from the camera module is a binary image with smoothed and polished contours. A contour is a list of points that represent a curve in an image. By setting a minimum size threshold, we removed contours with very small areas as those are probably unwanted noise. Polygon approximation is applied on the extracted contour to make it more suitable for shape analysis in the next step.

• Interface module: This module is responsible for translating the detected hand and fingers into functional inputs and interfacing with other applications. This module is responsible for translating the detected hand gestures into meaningful actions such as mouse clicks, virtual keyboard key input. These actions are then passed to the appropriate application based on the chosen output method such as passing mouse and virtual keyboard events. It is also able to simulate mouse and virtual keyboard events in Windows environment.

V. RESULTS

The proposed CAMShift algorithm with Kalman filter is implemented with OpenCV. The proposed algorithm is compared with two metrics accuracy and precision based on total number of a gesture actually displayed and total

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

number of the gesture predicted by the system correctly. The actual and the predicted gestures with different users are considered to construct confusion matrix. The results show that,

Measures	Web Camera						Leap Camera					
	1	2	3	4	5	Average	1	2	3	4	5	Average
Accuracy	0.70	0.60	0.60	0.60	0.70	0.64	0.80	0.90	0.90	0.60	0.80	0.80
Precision	0.75	0.67	0.71	0.67	0.60	0.68	0.80	1.00	0.80	0.60	0.83	0.806

Table 1

as calculated in Table 1, Leap Motion controller's accuracy, as well as precision, is more than Web-camera module for Human Computer Interaction system.

VI. CONCLUSION AND FUTURE WORK

The hand gesture based human computer interaction system provides a natural way to interact with a computer. The hand is first segmented by using skin colour information and then tracked using 'CAMShift' tracker with Kalman filter, then fingertips are located on the contour of the segmented hand and single gestures drawn from fingertips are recognized. Mouse events are generated to interact with the system. In future work, the hand segmentation is to be improved by using modified segmentation algorithm with more training.

REFERENCES

1. A. Ghotkar and K. Kharate, "Hand segmentation techniques to hand gesture recognition for natural human computer interaction," International Journal of Human Computer Interaction (IJHCI), vol. 3, no. 1, pp. 15–25, 2012.
2. A. Ghotkar, R. Khatal, S. Khupase, S. Asati, and M. Hadap, "Hand gesture recognition for Indian sign language," International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1–4, 2012.
3. Q. Wang, J. Cheng, and J. Pang, "A novel projector-camera interaction system with the fingertip," Journal of Image and Graphics, vol. 1, no. 2, pp. 80–84, 2013.
4. S. Akhalaq, H. Shah, A. Ahmed, I. Mohmood, and K. Khurshid, "Hand gesture based user interface for computer using a camera and projector," Signal and Image Processing Applications (ICSIPA), IEEE International Conference, pp. 168–173, 2011.
5. C. Chen, M. Zang, K. Qiu, and Z. Pan, "Real-time robust hand tracking based on CAMShift and motion velocity," Digital Home (ICDH), IEEE International Conference, pp. 20–24, 2014.
6. L. Baraldi, F. Paci, G. Sera, and L. Benini, "Gesture recognition in ego-centric videos using dense trajectories and hand segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 688–693, 2014.
7. P. Chien, Y. Miao, and J. Guo, "A 3D hand tracking and gesture control in complex environments," IEEE Conference Publications, pp. 1–4, 2015.
8. A. Dhote and S. Badwaik, "Hand tracking and gesture recognition," in IEEE International Conference on Pervasive Computing (ICPC), 2015.
9. P. Nanivadekar and V. Kulkarni, "Indian sign language recognition: Database creation, hand tracking and segmentation," in International Conference on Circuits, Systems, Communication and Information Technology Appliances (CSCITA), pp. 358-363, 2014.
10. J. Y. Lee and S. I. Yoo, "An elliptical boundary model for skin color detection," in Proc. of the 2002 International Conference on Imaging Science, Systems, and Technology, 2002.
11. A. Salhi and A. Y. Jammoussi, "Object tracking system using CAMShift, Meanshift and Kalman filter," in World Academy of Science, Engineering and Technology, Volume 64, pp.674-679, 2012.
12. A. M. Burns and B. Mazzarino, "Finger tracking methods using EyesWeb," Gesture Workshop, pp. 156–167, 2005.
13. D. Leonardo, M. Lizarazo, J. Antonio, and T. Borja, "Hand position tracking using a depth image from a RGB-d camera," Industrial Technology (ICIT), IEEE International Conference, pp. 1680–1687, 2015.
14. J. Zhuang, "Compressive tracking based on HOG and extended haar-like feature," Computer and Communications (ICCC), IEEE International Conference, pp. 326–331, 2016.
15. X. Zabulis, "Vision-Based Hand Gesture Recognition for Human-Computer Interaction," in the universal access handbook 34, pp. 30, 2009.
16. Q. Chen, "Real-time vision-based hand tracking and gesture recognition," PhD diss., University of Ottawa (Canada), 2008.
17. W. Huimin and Z. Xiaoshi, "Improved and efficient object tracking algorithm based on camshift," Computer Engineering and Applications, Vol.21(10), pp.4216-4220, 2007.
18. G. Xu, D. Zhao, Q. Zhou, and D. Huang, "Moving target tracking based on adaptive background subtraction and improved camshift algorithm," in Proc 2012 International Conference Audio Language Image Process, pp.919-924, 2012.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

19. Z. Li and G. Liu, "Target tracking based on Mean-shit and Kalman filter," Journal of Projectiles, Rocket, Missiles andGuidance,Vol. 28(1), pp.71-74, 2008.

BIOGRAPHY

Subhash Nevhal is an M.E Student in the Computer Engineering Department, Pune Institute of Computer Technology, Pune, Maharashtra, India. His research interests are Computer Vision, Robotics and Machine Learning.