International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1432-1440

Trajectory-based Cognitive Recognition of Dynamic Hand Gestures from Webcam Videos

Richa Golash[1], Yogendra Kumar Jain[2]

[1] *Ph.D., Student, Samrat Ashok Technological Institute, Vidisha, Madhya Pradesh, India.*

2 Professor, Department of Electronics & Instrumentation Engineering, Samrat Ashok Technological Institute, India.

ORCID ID: 0000-0001-5873-9594

Abstract

An online description of the hand and detection of the centroid of a hand movement are critical requirements to trace the dynamic hand gestures. The hand acquires very small area in the image frame, and due to its nonrigid nature, and random behavior in movement, the quality of images in the video are affected considerably, if the movement is recorded from low-resolution camera for e.g. webcam. The image quality further reduces by the real-time factors associated with the background and thus, hand detection and its localization become a challenge in webcam videos. These challenges compel researchers either to work with static hand postures or to use advance sensor-based cameras. In this paper we have proposed a novel method to recognize and plot the trajectory of dynamic hand gestures directly in true color videos acquired through webcam. The cognitive learning of Scale-Invariant Feature Transform (SIFT) features of active hand template in each consecutive frames of the video help in tracing the path of hand movement without background subtraction or involving segmentation process. The determination of active hand template using Chen-Vese model makes our method invariant to the hand posture used by the user to perform his hand gesture. To test the efficiency of the methodology we have generated three different real-time common scenarios for a user to perform his hand movement. The empirical results obtained in different experiments demonstrate that the approach can withstand the challenges associated with the detection and tracking dynamic hand gestures when recorded from a simple webcam.

Keywords: Computer Vision, Machine Learning, Visual Object Recognition, Active Contour, Feature Extraction, Scale Invariant Feature Transform (SIFT), Visual Object Tracking.

I. INTRODUCTION

Applications based on vision-based hand gesture recognition (v-HGR) have become one of the most active research areas among researchers. This field has become very wide and has opened an era of natural and contactless control of any general-purpose machine associated with home, offices, hospitals, etc. through hand movements [1-3]. Recently, this field has shown high potential, to create smart and user-friendly environment for all age group users.

Implementations of v-HGR has two main segments, first is static hand gesture recognition (s-HGR) and second is dynamic hand gesture recognition (d-HGR). The hand is a non-rigid object because it has very high degree of freedom and has random behavior in movement. Owing to these characteristic images of the hand region suffer from self-occlusion, blur etc. The background conditions also play significant role in challenging the implementations of HGR in real-time applications [4], [5]. Hence, researchers either work on static hand postures or use advanced sensor-based camera to reduce the complexities in designing real-time DHGR applications.

Broadly d-HGR has two main components in the algorithm. The first component is the detection of the region of interest (ROI). The second component is continuous localization of ROI i.e. moving hand region in each frame of the video. The robustness of both the stages is a crucial requirement in the complete design of d-HGR. Due to high flexibility of hand geometry, users freely use different hand postures to perform hand gestures and due to the default natural randomness in hand movement, user use different trajectory to perform gestures of same meaning. Therefore, it is highly unpractical to build the system of dynamic hand gesture recognition on a pretrained model. Another issue which greatly affect the system is the real-time factors associated with the environment for e.g. illumination variation, artifacts etc. These challenges are further enhanced when the video is acquired from a simple camera which can provide only RGB (2-D) information of the image. The 2-D data provided by the webcam is insufficient to derive both the spatial and the temporal information of a moving non-rigid object such as hand in real-time environment [6], [7].

Hand-gesture recognition and visual object tracking has garnered a wide recognition in literature and been reviewed by many scholars [8-12]. In [8] survey focused on segmentation, feature extraction and classification process in static hand gesture recognition. In [9] the survey mainly covers hand-image segmentation techniques based on color, depth data, and moment invariants. According to it, discrete moment invariant can produce better results in segmenting static hand poses. Authors in [10] have conducted their study on depth images. In their opinions, robust segmentation of hand region is a very important part in tracking. Meanwhile, segmentation based on skin-color suffers significantly from lighting changes, even if we use illumination-invariant color schemes. A detailed

comparative analysis including different taxonomies, the available techniques, application domain, commercial products and software related to HGR is discussed by [11]. According to [11] motion-based hand detection is not so much favored as hand contours are not clear, uncontrolled background is a prime challenge. The review presented in [12] has compared different methods using conventional RGB cameras with the ones that used RGB-D cameras. According to [12], recognition of hand gestures is challenging as the hand region has a relatively small area and have complex articulation. The recognition algorithm must have invariance with respect to the size, orientation, and the speed of the hand gestures [12].

According to many researchers RGB information is not sufficient to deduce spatial as well as temporal characteristic of the moving hand region in practical scenario. Researchers pre assumed many factors before they conduct their experiments. According to [13] hand movement is a non-linear movement and frequently exhibits change in speed and direction. Therefore, authors included the depth information to treat hand-region as a set of moving pixels. They took a reference vector for the initialization of hand location and added minimum eigen value in tracking using Adaptive Kalman Filter (AKF). In [14] both color and depth information are used for hand-detection. Model-based approach is used to eliminate the whole body before background subtraction. This process is repeated for every frame to perform gesture spotting. In [15], again depth-based hand-detection and then adaptive mean shift (DAM-Shift) tracking approach. According to them hand image does not contain many distinct characteristics as a face therefore in their approach they divided the image frame into blocks of depth data and created 10000 features for an image size of 20*20 pixels. Their tracking efficiency decreases when a hand is closer to a camera [15].

In [16] electrical signals produced by surface electromyography (SEMG) are added to the vision-based Leap Motion sensor signals to increase the gesture recognition rate. In [17] and [18], the depth data of hand region are directly used for finding different statistic of hand for example: the number of detected fingers, position of finger tips, hand orientation, and palm center. In this case only one or two hand posture are used successfully for hand movement.

In the current scenario, because of the high sensitivity of RGB images toward light and scale variation, simple cameras are mostly restricted to static HGR. Limited works have been carried out till now using simple camera in vision based dynamic HGR. Scholars of [4] proposed hand tracking using CAMshift algorithm. Initially, they used the Viola–Jones algorithm to detect hand region on a grey scale. To many techniques, including HSV color space thresholding, Suzuki's Border Following technique, and convex hull method are used to determine finger tips. In [19] design of static HGR is proposed using background subtraction and HSV color space. According to them, similar color objects or moving objects degrade the performance of their hand-detection scheme. Similarly, [20] detect static hand poses obtained from a web camera, using an HSV color space and a convex hull algorithm.

Mainly observed pattern for hand detection are background subtraction and frame difference. Combining these [21] proposed dynamic hand tracking using the KLT feature tracker. To make their system robust, they simultaneously produce frame difference for both color and grayscale images. According to [21] the number of features obtained in KLT tracker gradually reduces in the succeeding frames of the video, therefore to reduce the chance of losing track or chance of moving in the wrong direction they have added CAMshift algorithm.

Our work is influenced by the usefulness of discriminative features and concepts proposed by [22]. According to [22], the dominant local features such as SIFT and SURF are more useful in finding hand region in the current frame and subsequently capable to describe trajectory of hand motion, without performing any foreground and background segmentation. [23] also used SURF features to track hand motion in an uncontrolled background. In this SURF features of two consecutive frames are matched against their counterparts to detect moving hand location. As they used region-growing approach to predict the location of hand in every next frame, the number of features extracted for matching increases rapidly. The main limitation in [22] and [23] is that they have extracted local features of complete frame which are very in large. The matching of large number of features consume more time and thus they have performed the experiment with low frame rate (8-16 FPS) and small gesture length (approx. 40 frames). [24] also used SIFT features in developing HGR technique but the work mainly focuses on static images.

The literature survey of vision-based DHGR indicates that detection of moving area of interest and its tracking are challenging processes in many aspects and researchers have to add too many algorithms in order to improve their efficiency. Therefore, to detect the trajectory of dynamic hand gestures from a RGB videos acquired through webcam, require a reliable solution to detect the region of interest, and a robust methodology that can overcome issues involved in tracking of non-rigid objects in real-time background. This paper aims to give a simple yet reliable solution for detecting and recognizing the movement pattern of moving-hand in RGB videos captured using a simple camera. The distinguishing features of our approach are:

(i) For every data sequence, our technique determines the online shape of moving hand using the Chen-Vese model [25]. We refer this shape as unique online hand model or active hand template (AHT). Thus, our technique is invariant to the hand posture used by the user to perform hand gestures.

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1432-1440

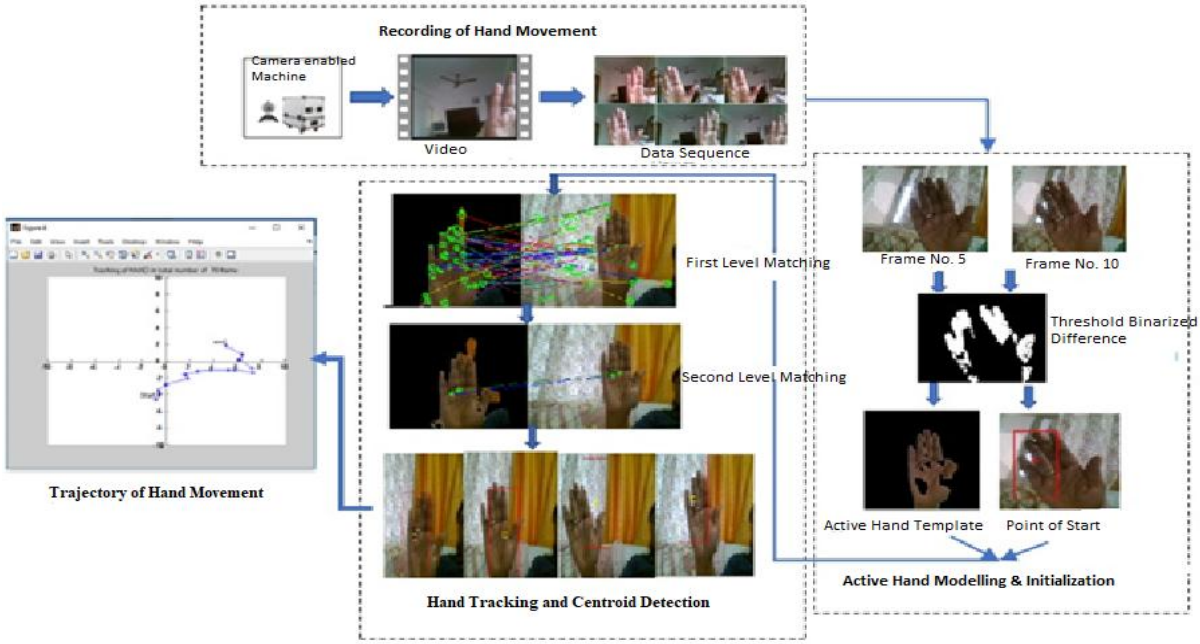


Fig. 1. The system architecture

(ii) In this method, we have directly localized the region of interest in each frame of video by using AHT, without performing background subtraction or segmentation process. The complete process has been accomplished directly on RGB images (which by nature are actually sensitive to illumination variations) using SIFT algorithm. Thus, our technique is invariant to background conditions.

II. THE SYSTEM ARCHITECTURE & ALGORITHM

In this section, we discuss the system architecture as shown in figure 1. The complete algorithm is divided into three stages after the movement is recorded and converted into frames:

II.I Active Hand Modeling and Initialization

Accurate hand region detection is essential for a reliable and successful tracking. Nowadays, applications have to be user-friendly and designed for all age groups. Thus, use of fixed parameters of hand shape and size while designing DHGR for real-time applications is not feasible. In the proposed method we have not used any fixed parameter, and whenever video is captured the method extract the unique hand posture used by the user. The process of hand modeling is divided in two parts: (i) Determination of online model of the target, known as active hand modelling, and (ii) Detection of location from where the target has initiated movement, known as position of start (POS).

To determine the online model, we produce pixel-wise intensity difference by subtracting two initial frames with a frame difference of five. While performing hand movement, the hand region is the most active and closest-moving area in front of the camera. Therefore, frame difference of five helps to determine objects with substantial displacement and avoid objects with negligible movement.

$$Temporal\ Diff: D_T = F_{10} \sim F_5 \quad (1)$$

The absolute frame difference D_T now consist of bimodal gray level distribution therefore automatic thresholding using Otsu method [26] is a better choice to divide the pixels of D_T into two classes C_1 and C_2. Here C_1 and C_2 corresponds to the foreground (objects of interest) and the background pixels. Using threshold t, D_T is converted to threshold binarized difference image' D_{OT} . Since hand is the closest moving object in front of camera, largest connected area is detected by using differential blob detector technique based on the Laplacian of Gaussian (LoG). Here we convolve D_{OT} by Gaussian Kernel $g(x,y,t) = \frac{1}{2\pi t}\, e^{-\frac{x^2+y^2}{2t}}$ to give scale space representation $L(x,y;t)$ of the image at certain scale t, equation (2) [27].

$$L(x,y;t) = D_{OT} * g(x,y,t) \quad (2)$$

$$\nabla^2_{norm} L = t(L_{xx} + L_{yy}) \quad (3)$$

$L(x,y;t)$ is further normalized to capture multilevel blob with automatic scale selection given by equation (3)

The point which is maxima or minima of $\nabla^2_{norm}L$ at both scale and space is our interest point $(\hat{x}, \hat{y})$ at scale $\hat{t}$ or position of start. The connected region around $(\hat{x}, \hat{y})$ is the largest blob or the region of hand in that frame [28].

$$(\hat{x}, \hat{y}, \hat{t}) = argmaxminlocal_{(x;y;t)}(\nabla^2_{norm}L(x; y; t)) \quad (4)$$

The area detected, consist of hand region and background. This area has convexity defect, to remove the extra part and to determine the model of the active moving target, we apply active contour model (ACM) proposed by Chan and Vese [25], [28]. The basic aim of Chan and Vese (CV) model [25] is to partition a given image into two regions that are likely to correspond object and background regions. This is done by embedding the object boundary using zero-level curve of a 3D level set function. Let Φ be a level set function, then, the Chan-Vese functional is given as equation (5) [25]

$$E^{CV}(c_1, c_2, \Phi) = \lambda_1 \int_{\Omega} (D_T - c_1)^2 H(\Phi)\, dx + \lambda_2 \int_{\Omega} (D_T - c_2)^2 \big(1 - H(\Phi)\big)\, dx + \mu \int_{\Omega} |\nabla H(\Phi)|\, dx \quad (5)$$

Where Φ is the level-set function defined as the region $\{(x, y) \in \Omega: \Phi(x, y) = 0\}$, when the zero-level set Φ comes to steady state it becomes the contours that separate the moving object from the background. $\lambda_1, \lambda_2 > 0$ *and* $\mu \geq 0$ are fixed parameters. The length parameter is interpreted as scale parameter and possibility of detecting smaller object increases by decreasing the value of μ. Keeping c_1, c_2 fixed and $\delta(t) = \frac{d}{dt}H(t)$ the zero-level set is computed using Euclidean distance transform of the binary threshold difference image D_{OT}. For each pixel in D_{OT}, the distance matrix assigns the distance between that pixel and the nearest nonzero pixel of D_{OT}. Through iteration the zero-level line of the evolving level set function Φ come close to object boundaries. Hence, we obtain a piecewise constant approximation of the moving hand.

Figure 2 illustrates step by step outcomes of active hand modelling and initialization stage. An experiment is performed in a real-time background where the subject is in a room with non-uniform illumination. Figure2(a) and 2(b) are frame number 5 and 10 respectively of the data sequence. Figure2 (c) is binarized threshold difference frame (D_{OT}). In figure2(d) level-set function for *the* D_{OT} frame is defined $\{(x, y) \in \Omega: \Phi(x, y) = 0\}$. In figure2(e) active contours of the hand region are shown in green color. Finally, figure 2(f) is the active hand template (AHT) for this data sequence. In figure 2(g) the red color box with red '+' sign is the location of the point of start (POS) for hand tracking and centroid detection stage.

II.II Hand Tracking and Centroid Detection

Tracking an object means finding object location in a complete frame by using some unique properties of the object and then tracing its centroid of movement in the whole data sequence. Successful searching of an object has two obstacles. The first one is natural; created by the object itself. For example, blur due to speed variation, scale variation due to change in the distance of object from camera, and occlusion due to non-rigid nature of the fingers. The second obstacle is created by the environment; effects of background objects, illumination variation, and default noises which occur when dynamic movement is captured. Continuous background subtraction and segmentation are not feasible for real-time tracking, as it is not possible to create background model for every sequence, also background subtraction is a repetitive process that make real-time tracking slow.

In the proposed method, instead of finding the exact location of the palm's center, the displacement of hand in between two consecutive frames are calculated as the displacement of dominant local feature (AcSIFT) of hand in those frames. We have developed a technique that can track hand movement in RGB images directly without converting images into gray or binary space. The two-level matching strategy of the SIFT algorithm, designed and described by David Lowe, is the key element in our tracking process [29]. It enables our technique to track complex hand movement in a real-time without being affected by illumination conditions or performing any segmentation actions. The local features detection technique identifies locations in an image that are invariant to image translation, scaling, and rotation, by using a staged filtering approach. These features have a high distinctiveness and better detection accuracy toward local image distortions, viewpoint change, and partial occlusion, and are helpful in real-time fast-tracking of a target [29],[30]. SIFT feature has both detector and descriptor. Each feature point detected is specified by four parameters: $f_i = \{p_i, \sigma_i, \varphi_i, gh_i, d\}$. where $p_i = (x_i, y_i)$ is the 2D position of SIFT key-point, σ_i is the scale, φ_i gradient orientation within the region and d is 128-dimensional descriptor of key point 'i'. To remove ambiguous features and to retain only stable key features we apply a condition on $D(X)$ where D is difference of gaussian function and $D(X)$ is given as

$$D(X) = D + \frac{1}{2}\frac{\partial D^T}{\partial X}X + \frac{1}{2}X\frac{\partial^2 D}{\partial X^2}X \quad (6)$$

where $D(X) < 0.8$. Let the active hand template (AHT) as shown in figure 2 (f) has 'm' SIFT key features given as $S_g = \{f_i\}^m$ where f_i is the feature vector at i^{th} location. $S_c = \{f_j\}^n$ are 'n' SIFT features in current frame where f_j is the SIFT at j^{th} location. Since area of AHT remain fixed throughout the experiment and it is small as compared to the frame size and it. Therefore, number of SIFT features in S_{AHT} are very less as compared to the number of features of whole frame and they remain fixed. It is unlike with the other methods which have used local features [22],[23]. The resultant stable key points of current frame are matched with the key features of AHT using nearest neighbour method as shown in equation (7), figure 3(b).

International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1432-1440

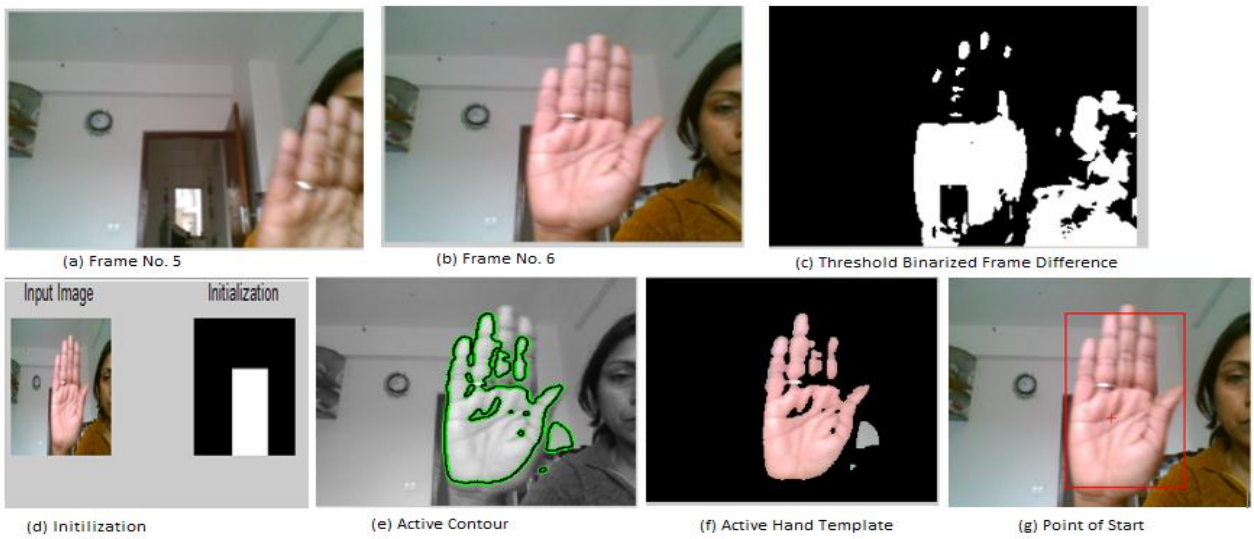


Fig. 2. Results of Stage I: Active Hand Modelling and Initialization (a) Frame No. 5 (b) Frame No.10 (c) D_OT Image (d) Active Contour Initialization (e) Active Contour (f) AHT (g) POS

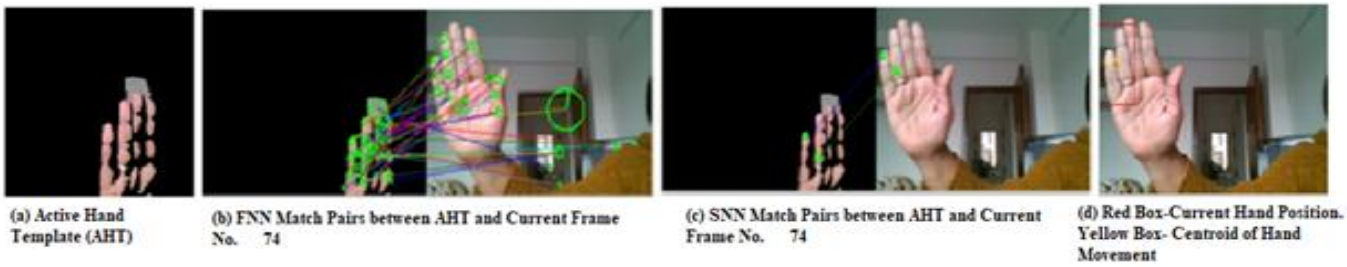


Fig. 3. Stage II: Hand Tracking and Centroid Detection (a) AHT (b) FNN pairs between AHT and current frame (c) SNN pairs between AHT and current frame (d) Centroid C_{HM} detection in the current frame.

The first nearest neighbor (FNN) is defined as the pair of key points with minimum sum of squared differences for the given descriptor vector.

$$distance\ (a_g, b_c) = \sqrt{\sum_{i=1}^{128}(a_i - b_i)} \qquad (7)$$

Where a_g and b_c are descriptor vectors of features in AHT and current frame respectively. Each feature of AHT are matched against their counterparts in current frame. To narrow down the numbers of first nearest neighbours we initiate the second level matching called second nearest neighbour (SNN). This is done by calculating the ratio between the second nearest neighbour distance (SNND) and the first nearest neighbour distance (FNND) of matched feature pair. For each descriptor $a_g$in the AHT, let the two nearest neighbours are b_c and e_c in the current frame. If value of the equation (8) is less than 0.8 then b_c is selected over e_c and vice-versa. This step gives features which are very close match of AHT key-points in the current frame. Out of all the close matches obtained in second level matching, we now determine the unique key-feature which has least distance among the pairs. The location of that key feature is selected as the centroid of hand movement C_{HM} in that current frame. A bounding box of size calculated in stage 1 (Active Hand Modeling and Initialization) is constructed taking C_{HM} as center.

$$\frac{distance(a_g,b_c)}{distance\ (a_g,e_c)} \qquad (8)$$

II. III Trajectory of Hand Movement

This is the final stage of tracking dynamic hand movement and the foundation stage for recognizing the pattern of hand movement. The inputs of this stage help the general-purpose machine understand and interpret hand gestures as commands. However, this stage has not been explained by scholars in a clear manner. But this stage is significant because the manner in which we collect and transform the C_{HM} of every frame of a particular data sequence, help in the design of classification stage and in turn it makes interpretation of visual hand commands fast and error free. In this method, we have used the concept of quadrant system of the Cartesian plane to transform the image frame into 2-D plane. The two-dimensional Cartesian system divides the plane of the frame into four equal regions called Quadrants. Each quadrant is bound by two half-axes, with the center in the middle of a frame. The translation of image frame axis to cartesian axis is done using equation (9) and (10):

$$X_c = (C_{HMx} - M_x)/n_x \qquad (9)$$

$$Y_c = (C_{HMy} - M_y)/n_y \quad (10)$$

Here M_x, M_y are parameters of image frame and n_x, n_y are normalization factors for X and Y axis. Figure 4(a) shows eight intermediate frames of a data sequence, here subject is moving his hand from bottom right to the top right of the frame. In each frame red box tells the location of hand in current frame and yellow box is the location of centroid of movement. Figure 4(b) is the collection and plot of C_{HM} in the Cartesian plane. The plot shows the start position (POS) and end position of the hand movement. The plot in 4(b) justifies the movement in 4(a).

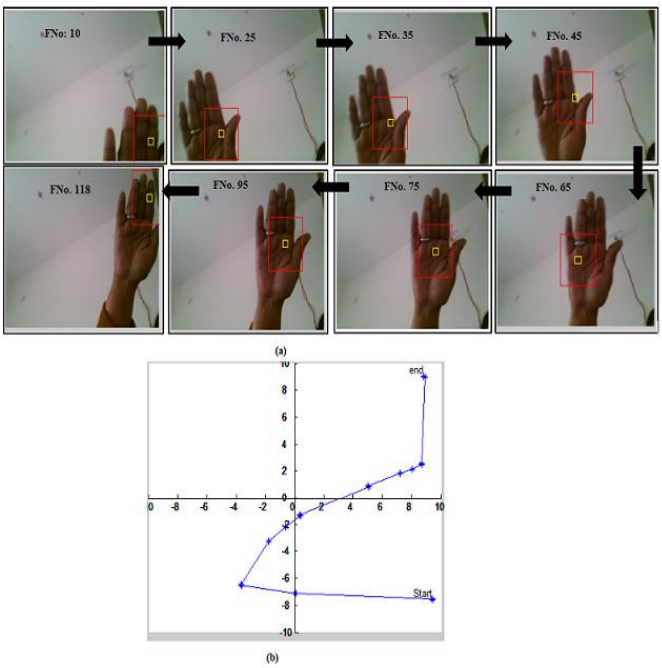


Fig. 4. Stage III: (a) Red box is the location of hand in current frame, yellow box is the centroid of hand movement. (b) Plotting of centroid of movement in Quadrant system

III. EXPERIMENTS AND RESULTS

The proposed technique is developed with a system having 2.16 GHz processor and 4.0 GB RAM. The system is enabled with Windows WDM compatible low-cost camera which can capture 15 frames per second. To test the proposed method, we have recorded approximately 100 videos of hand movement under three different cases (i) when different hand postures are used for movement (ii) illumination conditions are varied (iii) different background conditions are used. Number of frames in the recorded data sequence varies between 100-200 frames.

4.1 Case 1: Efficiency for different hand postures in movement: To test the efficiency of our online hand-detection scheme we have taken four most probable hand postures. We have collected 100 videos of hand movement, performed in real-time background with different illumination conditions. Fig. 5 (a), (b), (c) are the hand detection outcomes in proper illumination. Fig. 5 (d) is the outcome when user is also visible and position of hand is very close to camera. In fig. 5(e) illumination is less and scale of hand region is changed and fig. 5(f) is the outcome in poor illumination.

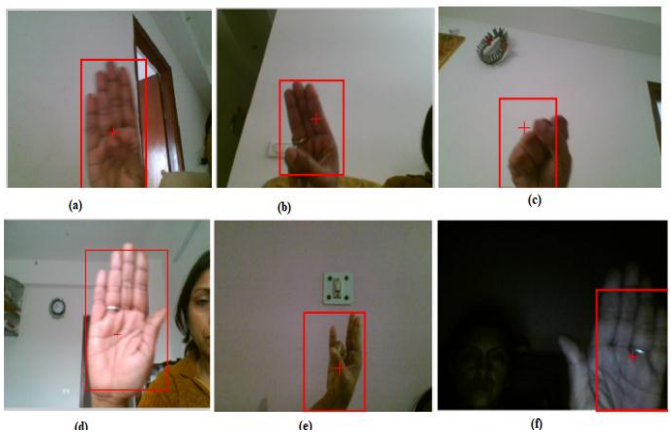


Fig. 5. POS detection in different data sequence captured different conditions.

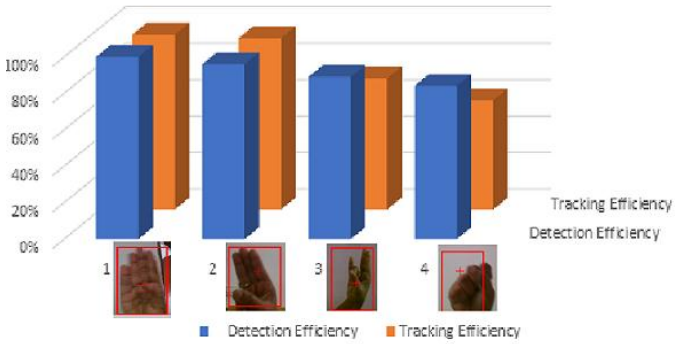


Fig. 6. Efficiency of hand detection and tracking of four different hand postures.

Fig. 6 compare the hand detection and tracking efficiency of four types of hand postures that are used during the hand movement. To test the efficiency of the proposed system 25 videos for each hand posture are recorded. It is observed that posture 1 has highest detection and tracking efficiency followed by posture 2. Posture 3 and posture 4 have good detection rate but their tracking efficiency is reduced.

4.2 Case 2: Efficiency of tracking in different illumination conditions:

To test tracking efficiency of proposed method we have selected four conditions as listed in table 1. The efficiency of the tracking scheme is calculated using equation (12).

$$Efficiency = \frac{True\ Detection}{(True + Missed\ )Detection} * 100 \quad (12)$$

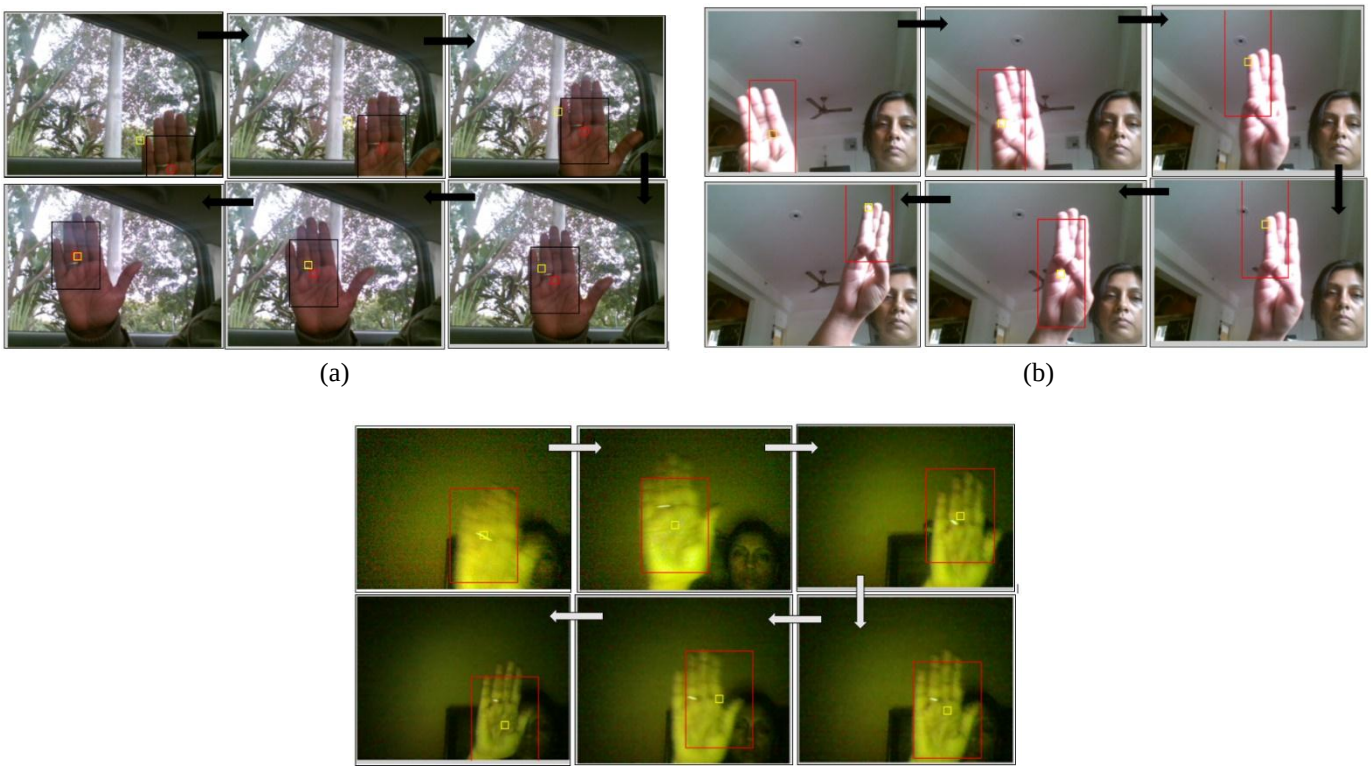
International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 6 (2020), pp. 1432-1440

Table 1. Comparison of the proposed method with Kalman algorithm.

Illumination Conditions	No. of frames used	Proposed Tracking Method		Tracking with Kalman Method	
		No. of false centroid detection	Efficiency	No. of false centroid detection	Efficiency
Uniform illumination	95	5	94.70%	20	78.94%
Poor illumination	95	8	91.50%	35	63.15%
Bright illumination	95	6	93.60%	21	77.8%
Dynamic illumination	95	7	92.63%	47	50.52%
Average Efficiency	-	-	93.10%	-	67.60%

The efficiency of the proposed technique is compared with the efficiency of Kalman algorithm in table 1. In this, we have taken 95 frames in every data sequence. The results in table 1 reflects that in case where illumination is changing dynamically the efficiency of Kalman Algorithm decrease to 50.52% in contrast to our proposed system which is 92.63%. The overall efficiency of our online region detection technique is 93.10 %.

4.3 Case 3: Efficiency in different background conditions: To test the robustness against real-time background conditions of the proposed method we have captured videos of hand movement in indoor as well as outdoor environment as shown in fig. 7. Fig. 7(a) demonstrates the tracking results in outdoor environment and fig. 7(b) demonstrates the results in indoor environment when subject is also in view and the data sequence in fig. 7 (c) is captured in very poor illumination condition.



(a) (b)

(c)

Fig. 7. (a) Tracking in outdoor environment (b) tracking in indoor environment.

VI. CONCLUSION

In this paper, we have proposed the online solution for hand-detection and reliable tracking in real-time RGB images. The empirical results of the experiments conducted in various conditions demonstrate that hand-detection using the active contour method in the initial stage overcomes the challenges of using any predefined geometrical values of handshape for the hand detection. The integration of the active hand template with SIFT features reduces the time of search in matching and recognizing the moving hand region throughout the video. The approach directly works on RGB video captured through 2-D webcam, without putting any constrain on background conditions or creating motion history images. The proposed method show invariance for hand shape or the trajectory of hand movement. The efficiency of the system is 93.10% with a normal camera. In the future, the integration of the proposed method and neural network algorithm has a scope to develop smart, manipulative, cost-efficient user-friendly natural interfaces.

REFERENCES

[1] Wachs, J.P.; Kölsch, M.; Stern, H; Edan, Y..2011. Vision-based hand-gesture applications. Commun ACM 2011; 54(2): 60-71.

[2] Mitra, S.; Acharya, T.. Gesture recognition: A survey. IEEE T Syst Man and Cy C 2007; 37: 311-324.

[3] Van den Bergh, M.; Carton, D.; De Nijs R.; et al.. Real-time 3D hand gesture interaction with a robot for understanding directions from humans. In Proceedings of the IEEE 2011 RO-MAN, Atlanta, GA, United States 31 July-3 August 2011, pp.357-362.

[4] Stergiopoulou, E.; Sgouropoulos, K.; Nikolaou, N.. Real time hand detection in a complex background. Eng Appl Artif Intel 2014; 35: 54-70.

[5] De Smedt Quentin, Hazem Wannous, and Jean-Philippe Vandeborre, "Skeleton-based dynamic hand gesture recognition," In 2016 IEEE conference on computer vision and pattern recognition workshops (CVPRW), IEEE, 2016, pp. 1206-1214.

[6] Kovalenko, M.; Antoshchuk, S.; Sieck, J.. RealTime Hand Tracking and Gesture Recognition Using Semantic-Probabilistic Network. In Proceedings of the 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, March 2014, pp. 269-274.

[7] Chakraborty BK, Sarma D, Bhuyan MK, MacDorman KF. Review of constraints on vision-based gesture recognition for human–computer interaction. IET Computer Vision. 2017 Nov 16; 12(1): 3-15.

[8] Khan, R.Z.; Ibraheem, N.A.. Hand gesture recognition: a literature review. International journal of Artificial Intelligence & Applications 2012, 3, 161.

[9] Yang, S.; P. Premaratne, P.; P. Vial, P.. Hand gesture recognition: An overview. In Proceedings of the 5th IEEE International Conference on Broadband Network & Multimedia Technology, Guilin, China, 2013, pp. 63-69.

[10] Suarez, J.; Murphy, R. R.. Hand gesture recognition with depth images: A review. In Proceedings of the RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, September 2012 pp. 411–417.

[11] Rautaray, S. S.; Agrawal, A.. Vision based hand gesture recognition for human computer interaction: a survey. Artif Intell Rev 2015; 43: 1-54.

[12] Pisharady, P.K; Saerbeck, M.. Recent methods and databases in vision-based hand gesture recognition: A review. Comput Vis Image Und 2015;141:152-65.

[13] Asaari, M.S.M.; Rosdi, B.A.; Suandi, S.A.. Adaptive Kalman Filter Incorporated Eigenhand (AKFIE) for real-time hand tracking system. Multimed Tools Appl 2015; 74: 9231-57.

[14] Xu, D.; Wu, X.; Chen, Y.L.; Xu, Y.. Online dynamic gesture recognition for human robot interaction. J Intell Robot Syst 2015; 77: .583-596.

[15] Joo, S.I.;Weon, S.H.; Choi, H.I.. Real-time depth-based hand detection and tracking. The Scientific World Journal 2014, Article ID 284827, 17 pages

[16] O. Kainz and F. Jakab, "Approach to hand tracking and gesture recognition based on depth-sensing cameras and EMG monitoring", Acta Informatica Pragensia, 2014; 13(1): 104-12.

[17] Marin, G.; Dominio, F.; Zanuttigh, P.. Hand gesture recognition with jointly calibrated leap motion and depth sensor. Multimed Tools Appl 2016; 75(22): 4991-15015.

[18] D. Zhao, Y. Liu and G. Li, "Skeleton-based Dynamic Hand Gesture Recognition using 3D depth data," Electronic Imaging, 2018; 2018 (18): pp. 1-8, 2018

[19] Chen Z.H., Kim J.T., Liang J., Zhang J. and Yuan Y.B., Real-time hand gesture recognition using finger segmentation, The Scientific World Journal 2014 (2014).

[20] Gurav, R.M.; Kadbe, P.K.. Real time finger tracking and contour detection for gesture recognition using OpenCV. In the Proceedings of the International Conference on Industrial Instrumentation and Control (ICIC) May 2015, pp. 974-977.

[21] Singha, J.; Roy, A.; Laskar, R.H.. Dynamic hand gesture recognition using vision-based approach for human–computer interaction. Neural Comput and Appl 2018; 29(4): 1129-41.

[22] Bao, J.; Song, A.; GuoY.. Dynamic hand gesture recognition based on SURF tracking. In Proceedings of the International Conference on Electric Information and Control Engineering, Wuhan, China, May 2011, pp. 338-341,.

[23] Yao Y.; Li, C.T.. Real-time hand gesture recognition for uncontrolled environments using adaptive SURF tracking and hidden conditional random fields. In Proceedings of the International Symposium on Visual Computing 29 Jul 2013; Berlin, Heidelberg: Springer, pp542-551.

[24] Mahmud, H.; Hasan,M. K. ; Tariq ,A.A.. Hand gesture recognition using SIFT features on depth image. In Proceedings. of the 9th International Conference on Advances in Computer-Human Interactions (ACHI) 2016, pp. 359-365.

[25] Getreuer, P.. Chan-vese segmentation. Image Processing On Line 2012; 2: 214-24.

[26] Al-Bayati, M.; El-Zaart, A.. Automatic thresholding techniques for optical images. Signal & Image Processing 2013; 4: 1.

[27] Lindeberg, T.. Scale selection properties of generalized scale-space interest point detectors. J Math Imaging Vis 2013; 46: 177-210.

[28] Huang, C.; Zeng, L.. An active contour model for the segmentation of images with intensity inhomogeneities and bias field estimation. PloS one, 2015, 10, Article ID. e0120399.

[29] Lowe,D.G.. Distinctive image features from scale-invariant keypoints. Int J Comput Vision 2004; 60: 91-110.

[30] Tuytelaars, T.; Mikolajczyk, K.. Local invariant feature detectors: a survey. Foundations and trends® in computer graphics and vision 2008; 3: 177-280.