International Journal of Advanced Research in Engineering and Technology (IJARET)
Volume 11, Issue 6, June 2020, pp. 338-348, Article ID: IJARET_11_06_030
Available online athttp://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=6
ISSN Print: 0976-6480 and ISSN Online: 0976-6499
DOI: 10.34218/IJARET.11.6.2020.030

ECONOMICAL AND USER-FRIENDLY DESIGN OF VISION-BASED NATURAL-USER INTERFACE VIA DYNAMIC HAND GESTURES

Richa Golash* and Yogendra Kumar Jain
Samrat Ashok Technological Institute,
Civil Lines, Vidisha, Madhya Pradesh, India
*Corresponding Author Email: golash.richa@gmail.com

ABSTRACT

A vision-based hand gesture recognition technology can bring a revolutionary and beneficial change in the field of human-machine interaction for elderly and special people living at home. Nevertheless, continuous detection and localization of the moving hand region in true-color images are strenuous tasks because the hand is a non-rigid object and occupies a small area in the whole frame. True-color images are also sensitive to light variation, camera-view, and background conditions. To ease the process of hand detection and tracking, researchers prefer advanced cameras equipped with costly sensors. This increases the overall cost of interfaces and also requires technical knowledge to operate them. The second issue in dynamic hand gesture recognition is the unpredictability of hand pose used by the user and the random behavior of the hand movement while performing the hand gesture. Therefore, the use of dynamic hand gestures in vision-based human-machine interaction is limited. The goal of this paper is to propose an economical, and user-friendly, technique for vision-based human-machine interaction via dynamic hand gestures that is user-friendly and affordable by all. The proposed technique can be integrated with any day-to-day machines, for example, washing machines, radio, fans, automated doors, etc.

Key words: Computer vision, human-machine interaction, visual object recognition, feature visual object tracking

Cite this Article: Richa Golash and Yogendra Kumar Jain, Economical and User-Friendly Design of Vision-Based Natural-User Interface Via Dynamic Hand Gestures, *International Journal of Advanced Research in Engineering and Technology*, 11(6), 2020, pp. 338-348.
http://www.iaeme.com/IJARET/issues.asp?JType=IJARET&VType=11&IType=6

1. INTRODUCTION

The hand gestures have always been a successful as non-verbal communicating medium among human beings. With the rapid development in computer vision, and pattern recognition

field, researchers are now focused to develop hand gestures as a natural and contactless modality of interaction between human and machines [1], [2]. The advantage of developing vision-based user interfaces is that these devices are user-friendly especially for senior and specially-abled people at home (figure 1), who cannot move frequently to operate home appliances.

Till now many natural user interfaces (NUIs) have been created using sign languages of hand. The user shows some static postures of hand and the machine interprets hand posture as one of command according to the training given, for example Bergh M. et al. [3], designed control for robot movement, Ren Z. et al. [4] developed a vision-based arithmetic computation tool system and Rock-Paper Scissor game using depth data of hand postures, Ohn-Bar E. et al. [5] designed a vision-based gestural interface to control a car infotainment system, etc. The common approach in the most of the techniques is that they used depth images obtained from advanced sensor-based cameras and focused to recognize static hand postures in their design.

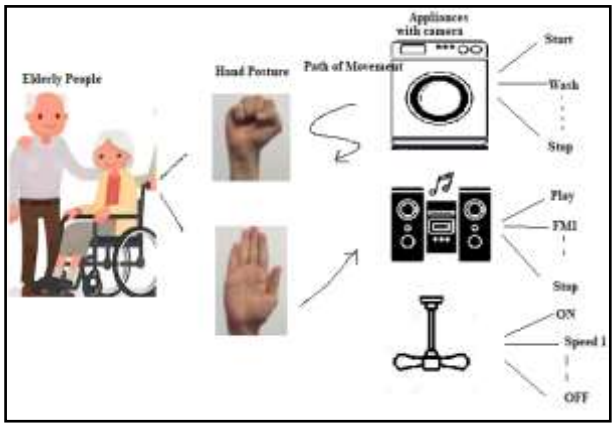

Figure 1. Model of vision-based control of home appliances

In contrast to static hand pose, dynamic hand gesture is a combination of spatial as well as temporal events. The process of dynamic hand gesture recognition involves simultaneous detection of hand pose and localization of the hand pose in a sequence of image frames. Visual tracking of hand motion is challenging because the hand has a non-rigid structure and does not move in a predefined manner. The moving hand region occupies a very small area in comparison with the whole frame and its images are affected by view angle, illumination variation or any type of occlusion, thus continuous temporal segmentation of moving hand in real-time background is difficult. According to Yang et al. [6], initial object description is a critical requirement for robust and efficient tracking and this description must cope with the variation in illumination, cluttered background, and occlusions. Different studies conducted in hand gesture recognition have acknowledged that the 3D model and the appearance-based are two main object representation techniques. Nowadays 3D anatomical description for example fingers position, palm center, hand orientation, etc. are easily determined, by using advanced sensor based-cameras such as Kinect, leap-motion, etc. [7]. The Kinect camera consist of an infrared laser projector with a monochrome CMOS sensor and the Leap motion camera consists of two cameras and three infrared LEDs. These special cameras have inbuilt software that is capable to provide position of the fingertips, 3D distance of the fingertips from the hand palm center, distances of the fingertip from the palm region plane, and the hand orientation, without doing much of computation work [8],[9],[10]. Nowadays it is becoming a trend to use the specialized advanced camera in the vision-based hand gesture recognition field to ease the initial process of hand detection and tracking. But this approach increases the

overall cost of the design of the natural user interface and also requires skilled people to operate these interfaces. Maintenance of advanced sensor-based user interfaces also become a costly affair. Hence, in vision-based human-machine interfaces, where affordability and usability are prime important factors in design, embedding of the specialized camera is not an economical and intelligent selection. This paper aims to propose a low-cost design of a vision-based natural user interface via dynamic hand gestures. Since the methodology utilizes webcam which is a familiar camera among all age group users, therefore this vision-based user interface is easy to understand and trouble-free to use.

A common observed pattern in the techniques, using true-color images (also known as RGB images) is either they focus for static hand postures detection [11], [12] or if motion is concerned then they have used small length gestures with constrain in background [13-15]. The reason being is that, images of hand movement are highly sensitive for camera viewpoint, background conditions. In addition to it, the hand does not move with a fixed speed, its boundaries and skin areas are quite uneven, hence edges of the moving hand region are not clear The problem is further enhanced if the experiment is recorded from an average quality camera, then segmentation and background subtraction also become a challenge. In comparison with global features, local features have high precision with respect to spatial information of hand and can give better results in motion tracking. Since the segmentation of moving hand is largely affected by skin color and a high degree of freedom in hand geometry, therefore local features are a good alternate for hand tracking in a real-time environment without detecting and segmenting the hand region [16- 18]. In [19], [20], Lucas-Kanade tracking algorithm is used to track the hand region in RGB images. The Lucas-Kanade-Tomasi (KLT) tracking algorithm works on the principle of brightness constancy assumption, therefore feature points decreases in large intraframe motion and tracker loses the track.

One very important issue which is missed or not been discussed more in the design of NUI is the development of semantic between trajectory and command. The main target in the design of semantic is that it should be invariant to hand postures and invariant to the subsequence path of gestures with same meaning. Due to the non-rigid nature of hand skeleton structure and random behavior in movement, researchers prefer static hand gesture recognition [4], [5], [12], [21] instead of dynamic hand gesture recognition. Tran D. et al. [22] proposed human-machine interaction via finger-tips tracking of seven hand contours extracted using Kinect V2. The sampling gesture was confined to 20-45 frames collected at a speed of 30 frames per second. Zeng J. et al. [23] developed NUI based on hand gestures for special people. They used a multi-cue system with frame-based motion history image but the motion was limited to the abduction and adduction process of thumb and three fingers in a black fixed background. Some scholars for e.g. Kılıboz N. et al. [24] and Grif H.S. et al. [25] even used external hardware attached to hand to record the hand movement.

Comprehensively analyzing the literature of HGR, highlights two major issues that limit the use of a low-cost camera in the design of vision-based user interface via dynamic hand gestures are:

(i) Hand movement is a spatio-temporal activity. Edges of the hand are not very clear and movement inherits a frequent change in camera view and scale. Therefore, continuous detection of the moving hand region in true color videos becomes difficult.

(ii) The user uses different hand postures and different paths to deliver the same messages. Area of some hand postures are very small and some postures are greatly affected by self-occlusion, that may lead to lose of trajectory. Therefore, segmentation, tracking and the semantic development is a rigorous task in the application based on dynamic hand gesture recognition.

we can derive that object description, continuous detection and finally, lexicon building are three important building blocks in the design of vision-based NUI via dynamic hand gestures. In this paper, our focus is to develop these three building blocks using a low-cost camera such that the designed technique is economical as well as user-friendly. The proposed methodology resolves issues of continuous hand detection and tracking in webcam images by using Active Scale Invariant Feature Transform (Active-SIFT) features of hand template. This approach avoids the unnecessary computation involved in matching a large number of features and then pruning ambiguous features. Localization of the dominant movement direction using the two-level comprehensive matching strategy of the SIFT algorithm enables us to describe hand trajectory without background subtraction and segmentation process. The second unique feature of the proposed method is that the technique is invariant to hand posture used by the user and is also the invariant trajectory of movement. In this method, we have developed eight trajectory-based commands (as shown in figure 3) for vision-based smart and natural interaction with any home appliances for e.g. washing machine, music system, fans, door, etc.

2. ARCHITECTURE OF PROPOSED SYSTEM AND ALGORITHM

Figure 2 illustrates the systematic working of the proposed methodology. It is divided into three modules as follows:

2.1. Module I: Hand Modelling and Initialization

Optimal description of moving hand region and detection of the start point of gesture play key roles in reliable and successful tracking in DHGR. Initially, we produce pixel-wise intensity difference D_T (equation (1)) by subtracting frame numbered 5th and 10th. Since we have captured video in real-time background, therefore images suffer from noises due to surroundings. D_T does not give a clear picture of moving objects and produces intraclass variance. Therefore, a unique threshold (calculated for each video) is applied on D_T to get 'Binarized Difference Image' D_{OT} [26]. D_{OT} image consists of only white and black pixels, where white pixels represent moving objects and black pixels stationary objects. Since the hand is the closest moving object therefore, we detect the largest connected area by using differential blob detector technique based on the Laplacian of Gaussian (LoG) [27]. D_{OT} is convolve 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)).

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

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

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

$L(x,y;t)$ is normalized to capture multilevel blob with automatic scale selection (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 [27].

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

In module I, we have two outcomes first is the determination of the active region of hand template (HT) and second is fixing of point of start (POS) of gesture. The active region of the hand template exhibits the hand posture used by the user to perform hand movement. We

have divided the image frame in quadrants and POS displays the quadrant from which the subject has started his gesture.

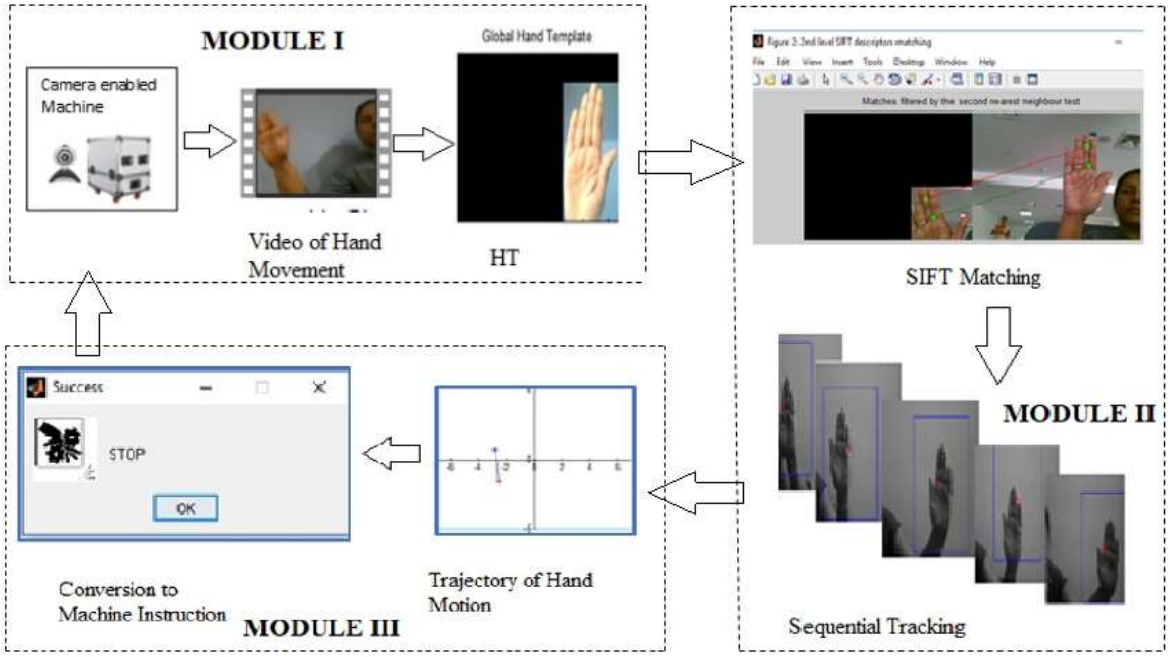

Figure 2. Architecture of the proposed system

2.2 Module II: Motion Modeling

Motion Modeling means determining the movement of hand gesture from start till the end and tracing the path in a meaningful manner. We have used the Scale-Invariant Feature Transform (SIFT) Algorithm designed and described by David Lowe in our tracking process [28]. Because SIFT 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 the target [28], [29]. SIFT algorithm describes the local characteristic of the image, 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 keypoint, σ_i is the scale, φ_i gradient orientation within the region and d is a 128-dimensional descriptor of the key point i. The unique attribute of our method is that to locate the moving hand region throughout the data stream we use SIFT features of HT frame only, which is the active region in the frame. We call the features of the HT template an Active SIFT (AcSIFT) features that are very small in numbers as compared to the whole frame. This approach reduces the time of computation occupied in matching SIFT features of one complete frame with the other frame [17]. It also avoids the unnecessary stages of region growing and pruning of ambiguous features [18].

Let there are m numbers of AcSIFT key features in HT frame, given as $S_g = \{f_i\}^m$, where f_i is the feature vector at i^{th} location. Let $S_c = \{f_j\}^n$ are n numbers of SIFT features in the current frame, where f_j is the SIFT feature at j^{th} location. We use the best-bin-first search method that identify the nearest neighbors of HT features with current frame features. The First Nearest Neighbors (FNN) are defined as the pairs of key points with a minimum sum of squared differences for the given descriptor vector [28].

$$distance\ (a_g, b_c) = \sqrt{\sum_{i=1}^{128}(a_i - b_i)} \qquad (5)$$

where a_g and b_c are descriptor vector of features in HT and current frame respectively. In the first phase of matching each feature of HT have multiple match pairs in the current frame. We observe that there many false and ambiguous or multiple match pairs between them. Therefore, to obtain the correct nearest match pairs we determine the second closest matching called Second Nearest Neighbor (SNN). This is done by calculating the ratio between the first nearest neighbor distance (FNND) of HT features with current frame features to the second nearest neighbor distance (SNND) of HT features with current frame features.

$$\frac{distance(a_G, b_T)}{distance\ (a_G, c_T)} \quad (6)$$

2.3. Module III: Machine Learning Algorithm

Module III is the final module in the design of NUI, here semantic is developed between hand gesture and machine command. The manner in which we plot the hand movement of a particular data sequence is the essence of generating a machine learning (ML) algorithm. The efficiency of the ML algorithm help machine to interpret the visual hand command easily and, in a time-efficient manner. In our prototype we have designed eight visible commands: START, Inst2 (Forward 2), Inst3 (forward 3), Inst4 (forward 4), Inst5 (forward 5), Inst6 (forward 6), Inst7 (forward 7), STOP. These commands are developed keeping in mind any general-purpose machine, especially used in home environment, and can be operated with 6-10 commands. The hand gesture is visualized as a trajectory, starting in a Cartesian quadrant and ending in a Cartesian quadrant.

There are two sets of instructions: Set A, Set B, depending on the start point of the hand gesture. Set A consists of instructions that are right initiated (Quadrant 4) and Set B consists of instructions that are left initiated (Quadrant 3). The figure 3 illustrates the semantic between hand movement and command. In the proposed method, the machine learning algorithm is built on the Modified Back Propagation of Artificial Neural Network (BP-ANN) using SIFT features. Back propagation (BP) is a supervised training procedure for 'feed-forward' neural networks. It works on minimizing the cost function of the network using delta rule or gradient descent method. The value of the weights with which we obtain minimum the cost function is considered to be the solution for the given learning problem. In traditional BP, the step size is fixed and hence could not optimize the multi-dimensional cost function. Also, the performance parameter is highly dependent on the value of the learning rate parameter δ. Hence convergence is very slow and increases overall learning time. Therefore, to counterbalance these two major problems BP is modified using the momentum term and adaptive learning rate. The updated weight value at any node is given as (equation (7)):

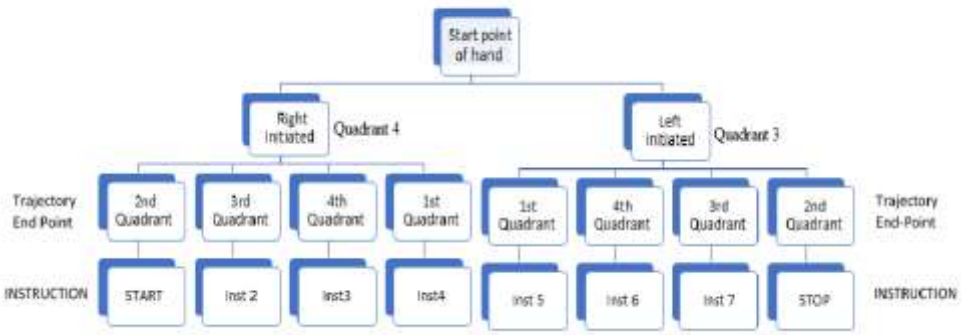

Figure 3. Semantic between hand movement and machine command.

$$\Delta w_{ij}(t) = \eta \delta_j a_i + m \Delta w_{ij}(t-1) \quad (7)$$

The momentum factor ($0 < m < 1$) increases the previous weight by a fraction and η adaptive learning rate help to learn the characteristic of the cost function. If error function decreases then the learning rate is η increased between 1% to 5%. If there is an increase in error function than the learning rate is decreased sharply. Thus, the value of the learning rate is continuously changed to adapt large value for fast learning or making small for its final, effective movement towards the real descent direction [30], [31].

3. EXPERIMENTAL RESULTS

The prototype of the methodology is developed with a system having a 2.16 GHz processor and 4.0 GB RAM. The system is enabled with Windows WDM compatible low-cost camera having 1280 x 720 pixels image resolution. We have captured 15 frames per second and created our own dataset which contains videos of moving hand, captured in different conditions, for example, lighting condition, object shape, motion smoothness of object movement, occlusion, the complexity of the background. The number of frames in video sequences varies between 100-200 frames. Five Subjects of 3 different age groups: two kids (age 10-16 year), two adults (age 20-40 years) and one senior (age 65 years) are selected to perform hand movement in 4 different styles, keeping in mind, start and end quadrants. Each instruction is repeated in 4 different ways by every user, so a total of (5×4=20) 20 test data sequences are generated for each instruction. For eight instructions, total 160 test data sequences and 16000 frames (160×100=16000) are created and tested to analyze the proposed system performance. The efficiency of the proposed method is analyzed on the basis of following parameters:

- Efficiency in extracting hand model and tracking the motion.
- Efficiency in the interpretation of machine command.

For the evaluation of hand detection and tracking efficiency, we have selected eight different real-time scenarios consisting of six indoor and two outdoor. we have captured sample gestures of variable length ranging from 50 frames to 200 frames and distance between hand and camera is also varied. Figure 4 shows the objective analysis of the hand detection scheme of the proposed method. We have performed experiments in two courses: first, with different postures and second, in different environments. Figure 5 illustrates the efficiency of tracking in different background calculated using equation (8). The sample tracking of two indoor and one outdoor sequence is demonstrated in figure 6. The results illustrate that the proposed algorithm is successful in tracking hand of different postures at different scales.

$$Tracking\ Efficiency = \frac{No.of\ correctly\ tracked\ frames}{Total\ No.of\ frames\ in\ a\ video} \times 100 \quad (8)$$

To test the instruction interpretation efficiency of the system we have collected data for each instruction. With the help of equation (9) we find that the accuracy of interpretation of trajectory to instruction as shown in table 1. We have captured approximately 20 real-time video sequences of each instruction. The efficiency of start and forward 5 instructions efficiency is 99% and forward 2, forward 3, forward 6, forward 7 are converted with an accuracy of 90%, and instruction forward 4 and Stop are converted with an accuracy of 85%. The average efficiency of the proposed vision-based dynamic interaction system 91%. Figure 7 demonstrates the complete working of the prototype of the proposed method.

$$Command\ Interpretation\ Accuracy = \frac{\text{No. of Correct Conversions}}{Total\ Sample} \times 100 \quad (9)$$

Figure 4. Some examples of Hand template detection (Module I).

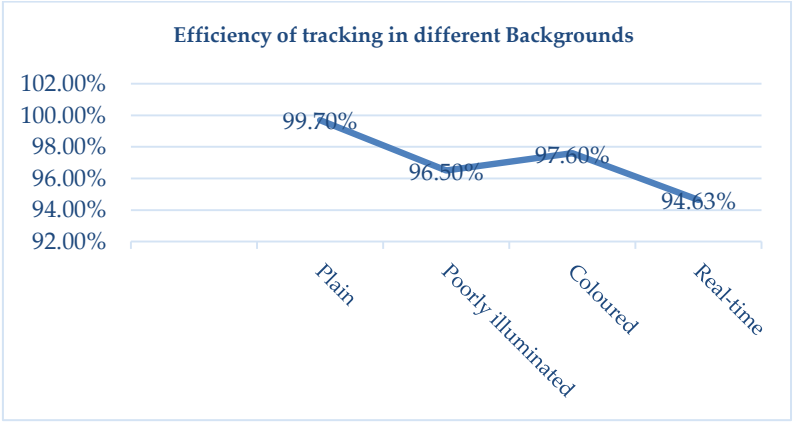

Figure 5. Efficiency of Tracking of hand movement in different background.

4. CONCLUSION

The empirical results of the proposed method are encouraging in the direction of economical and user-friendly design of natural user-interface for home appliances, for example radio, fans, washing machine. The determination of HT and matching SIFT features of HT with the current frame strategically, helps to trace the gesture path without any segmentation or any requirement of motion history images. This approach makes our method invariant to hand postures, and background conditions with an efficiency of 97.3 %. The integration of SIFT based centroid with Artificial Neural Network for tracing the path and classification of hand gesture to the machine command, make the method invariant to subsequence path adopted by different user. The technique is simple in its design and economical in cost, therefore it can be easily integrated with any day-to-day machine with affordable cost. This technique is first in its type where simple camera is used for the development of smart and manipulative user-friendly natural interface via dynamic hand gestures. In future, the integration of SIFT with deep Neural Network can improve the efficiency and the number of commands it can handle.

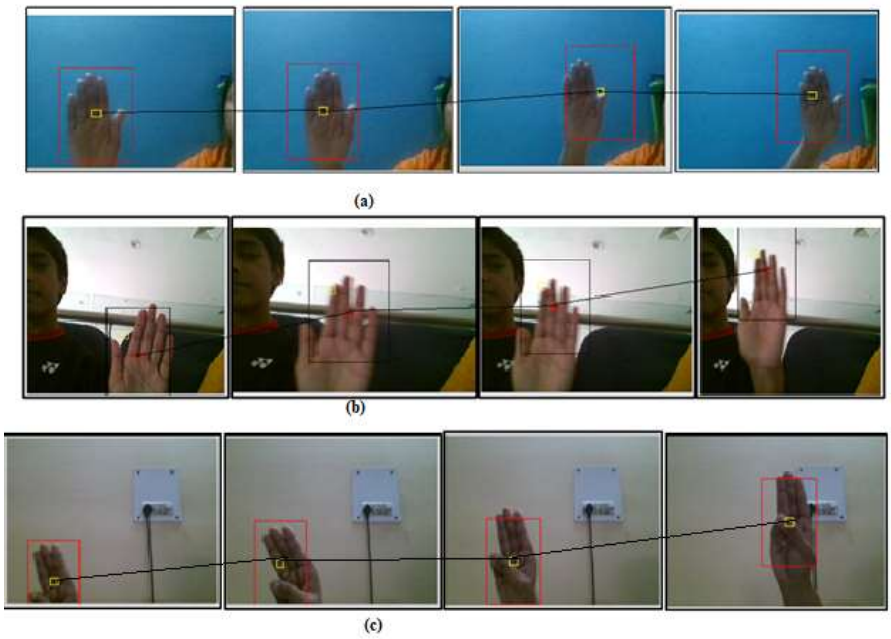

Figure 6. Tracking of Hand movement in three different data sequences (2 indoor and one outdoor), black line shows the trajectory of hand (Module II).

Table 1. Confusion matrix in interpretation of hand trajectories to instructions

Instruction	START	Inst. 2	Inst. 3	Inst. 4	Inst. 5	Inst. 6	Inst. 7	STOP
Start	19	1	-	-	-	-	-	
Inst. 2	1	18	1	-	-	-	-	
Inst. 3		1	18	1				
Inst. 4	-	-	3	17	-	-	-	
Inst. 5	-	-	-	-	19	1	-	
Inst. 6	-	-	-	-	1	18	1	
Inst. 7	-	-	-	-	-	1	18	1
Stop	-	-	-	-	-	-	3	17

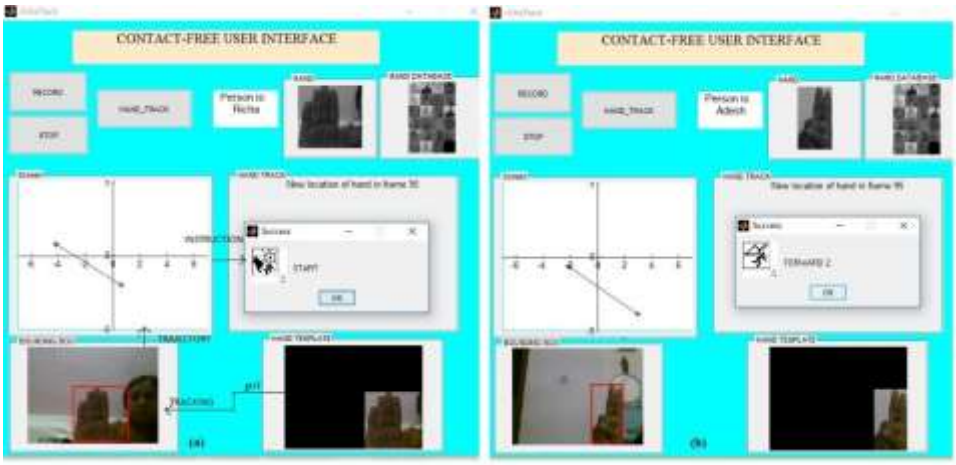

Figure 7. Execution of vision-based command (a) START command (b) Forward 2 command of Set A.

REFERENCES

[1] Rautaray Siddharth S., Anupam Agrawal, (2015) Vision based hand gesture recognition for human computer interaction: a survey, *Artificial intelligence review*, 43, 1, pp. 1-54.

[2] Pisharady Pramod Kumar, and Martin Saerbeck, (2015) Recent methods and databases in vision-based hand gesture recognition: A review, *Computer Vision and Image Understanding*, 141, pp.152-165.

[3] Van den Bergh, Michael, et al., (2011) Real-time 3D hand gesture interaction with a robot for understanding directions from humans, In 2011 Ro-Man, pp. 357-362. IEEE.

[4] Ren Zhou, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang, (2013) Robust part-based hand gesture recognition using kinect sensor, *IEEE transactions on multimedia*, 15, 5, pp.1110-1120.

[5] Ohn-Bar Eshed, and Mohan Manubhai Trivedi, (2014) Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations, *IEEE transactions on intelligent transportation systems*, 15, 6, pp. 2368-2377.

[6] Yang, Hanxuan, Ling Shao, Feng Zheng, Liang Wang, and Zhan Song, Recent advances and trends in visual tracking: A review, *Neurocomputing*, 74, 18, pp. 3823-3831, 2011.

[7] Suarez Jesus, and Robin R. Murphy, (2012) Hand gesture recognition with depth images: A review, In 2012 IEEE RO-MAN: the 21st IEEE international symposium on robot and human interactive communication, pp. 411-417. IEEE.

[8] Marin Giulio, Fabio Dominio, and Pietro Zanuttigh, (2016) Hand gesture recognition with jointly calibrated leap motion and depth sensor, *Multimedia Tools and Applications*, 75, 22, pp. 14991-15015.

[9] Zhao Dan, Yue Liu, and Guangchuan Li, (2018) Skeleton-based dynamic hand gesture recognition using 3d depth data, *Electronic Imaging*, 2018, 18, pp. 461-1.

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

[11] Chen Zhi-hua, Jung-Tae Kim, Jianning Liang, Jing Zhang, and Yu-Bo Yuan, (2014).Real-time hand gesture recognition using finger segmentation, *The Scientific World Journal* 2014

[12] Simion Georgiana, Ciprian David, Vasile Gui, and Cătălin-Daniel Căleanu, (2016) Fingertip-based real time tracking and gesture recognition for natural user interfaces, *Acta Polytechnica Hungarica*, 13, 5, pp. 189-204.

[13] Al Ayubi Shalahudin, Dodi Wisaksono Sudiharto, Erwid Musthofa Jadied, and Endro Aryanto, (2019) The Prototype of Hand Gesture Recognition for Elderly People to Control Connected Home Devices, In Journal of Physics: Conference Series, 1201, 1, p. 012042. IOP Publishing.

[14] Asaari Mohd Shahrimie, Mohd Bakhtiar Affendi Rosdi, and Shahrel Azmin Suandi, (2015) Adaptive Kalman Filter Incorporated Eigenhand (AKFIE) for real-time hand tracking system, *Multimedia Tools and Applications*, 74, 21, pp. 9231-9257.

[15] Ait Abdelali, Hamd, Fedwa Essannouni, Leila Essannouni, and Driss Aboutajdine, (2016). An adaptive object tracking using Kalman filter and probability product kernel, *Modelling and Simulation in Engineering* 2016

[16] He, Wei, Takayoshi Yamashita, Hongtao Lu, and Shihong Lao, (2009) Surf tracking, In 2009 IEEE 12th International Conference on Computer Vision, pp. 1586-1592, IEEE.

[17] Bao Jiatong, Aiguo Song, Yan Guo, and Hongru Tang, (2011) Dynamic hand gesture recognition based on SURF tracking, In 2011 International Conference on Electric Information and Control Engineering, pp. 338-341, IEEE

[18] Yao Yi, and Chang-Tsun Li, (2013) Real-time hand gesture recognition for uncontrolled environments using adaptive SURF tracking and hidden conditional random fields, In International Symposium on Visual Computing, pp. 542-551. Springer, Berlin, Heidelberg.

[19] Premaratne Prashan, Sabooh Ajaz, and Malin Premaratne, (2013) Hand gesture tracking and recognition system using Lucas–Kanade algorithms for control of consumer electronics, *Neurocomputing*, 116, pp. 242-249.

[20] Singha Joyeeta, Amarjit Roy, and Rabul Hussain Laskar, (2018) Dynamic hand gesture recognition using vision-based approach for human–computer interaction, *Neural Computing and Applications*, 29, 4, pp.1129-1141.

[21] Dinh Dong-Luong, Jeong Tai Kim, and Tae-Seong Kim, (2014) Hand gesture recognition and interface via a depth imaging sensor for smart home appliances, *Energy Procedia*, 62, pp. 576-582.

[22] Tran Dinh-Son, Ngoc-Huynh Ho, Hyung-Jeong Yang, Eu-Tteum Baek, Soo-Hyung Kim, and Gueesang Lee, (2020) Real-Time Hand Gesture Spotting and Recognition Using RGB-D Camera and 3D Convolutional Neural Network, *Applied Sciences*, 10, 2, pp. 722.

[23] Zeng Jinhua, Fang Wang, and Yaoru Sun, (2015) A natural hand gesture system for people with brachial plexus injuries, *Computing and Informatics*, 34, 2, pp. 367-382.

[24] Kılıboz Nurettin Çağrı, and Uğur Güdükbay, (2015) A hand gesture recognition technique for human–computer interaction, *Journal of Visual Communication and Image Representation*, 28, pp. 97-104.

[25] Grif, Horatiu-Stefan, and Cornel Cristian Farcas, (2016) Mouse cursor control system based on hand gesture, *Procedia Technology*, 22, pp. 657-661.

[26] Al-Bayati Moumena, and Ali El-Zaart, (2013) Automatic thresholding techniques for optical images, *Signal & Image Processing*, 4, 3, pp. 1.

[27] Lindeberg Tony, (2013) Scale selection properties of generalized scale-space interest point detectors, *Journal of Mathematical Imaging and vision*, 46, 2, pp.177-210.

[28] Lowe David G., (2004) Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, 60, 2, pp. 91-110.

[29] Tuytelaars Tinne, and Krystian Mikolajczyk, (2008) Local invariant feature detectors: a survey, *Foundations and trends® in computer graphics and vision*, 3, 3, pp.177-280.

[30] Moreira M., Fiesler E., (1995) Neural networks with adaptive learning rate and momentum terms, *Idiap*.

[31] Rojas R., (1996) Fast Learning Algorithms in Neural Networks, *Springer* pp.183-225.