

# GIFT: Gesture-Based Interaction by Fingers Tracking, an Interaction Technique for Virtual Environment

Muhammad Raees\*, Sehat Ullah

Department of Computer Science and IT, University of Malakand, KPK (Pakistan)

Received 28 October 2018 | Accepted 31 December 2018 | Published 25 January 2018



## ABSTRACT

Three Dimensional (3D) interaction is the plausible human interaction inside a Virtual Environment (VE). The rise of the Virtual Reality (VR) applications in various domains demands for a feasible 3D interface. Ensuring immersivity in a virtual space, this paper presents an interaction technique where manipulation is performed by the perceptive gestures of the two dominant fingers; thumb and index. The two fingertip-thimbles made of paper are used to trace states and positions of the fingers by an ordinary camera. Based on the positions of the fingers, the basic interaction tasks; selection, scaling, rotation, translation and navigation are performed by intuitive gestures of the fingers. Without keeping a gestural database, the features-free detection of the fingers guarantees speedier interactions. Moreover, the system is user-independent and depends neither on the size nor on the color of the users' hand. With a case-study project; Interactions by the Gestures of Fingers (IGF) the technique is implemented for evaluation. The IGF application traces gestures of the fingers using the libraries of OpenCV at the back-end. At the front-end, the objects of the VE are rendered accordingly using the Open Graphics Library; OpenGL. The system is assessed in a moderate lighting condition by a group of 15 users. Furthermore, usability of the technique is investigated in games. Outcomes of the evaluations revealed that the approach is suitable for VR applications both in terms of cost and accuracy.

## KEYWORDS

Gesture-based Interactions, Virtual Environments, Human Computer Interfaces, Image Processing.

DOI: 10.9781/ijimai.2019.01.002

## I. INTRODUCTION

VIRTUAL Reality is emerging in various domains of robotics [1-3], computer engineering [4,5], physical sciences, health-related issues [6-8], natural sciences [9] and industrial academic [10]. To ensure the feelings of being there while interacting with a 3D virtual space, an interface with a high degree of immersivity is required. Interaction via keyboard and mouse lags far behind to properly engross VR users [11]. The gestures of hands and/or fingers are used as the courier of feelings and thoughts in daily life. As gestures have meaning in real-world, hand gestures can be used for distinct interactions. Hence, with the gesture-based interactions, a VR interface can be made natural and perceptual [12].

Different interaction techniques have been proposed for acoustic mechanical and magnetic devices to make human-computer dialogue realistic. The complex setup of the devices make them a rare choice for interactions [13]. However, due to the recent developments in image processing, RGB and depth sensor cameras are becoming the prominent alternatives [14, 15]. Various state-of-the-art techniques have been proposed for HCI, based on RGB camera [16-18] and depth sensors [19-22]. Besides hand tracking, vision-based systems have been proposed for face tracking [23-25], action tracking [26-28] and gait tracking [29, 30]. Although, the outcomes of the contemporary research works favour gesture-based interfaces for interactions [31, 32], such systems are fairly more error-prone [33].

However, the errors and complexities may be minimized if simple

gestures are used for interactions [34].

This research work aims to introduce a low-cost 3D interaction technique based on the positions of the fingers in the dynamic image streams. With the simple and perceptive gestures of the fingers, a 3D object is selected and/or manipulated. At the detection of the finger-tips, a Virtual Hand (VH), indicating the user's position in the VE, is activated. The VH moves freely along x, y and z-axis with the movements of the hand. Navigation and panning are performed by the feasible movements of the hand. As bare-handed gestures are critical [35] and the recognition accuracy of gestures can be improved with coloured-markers [35], therefore the two contrast coloured markers are used for the robust and reliable detection of the fingers. Instead of extracting features from an entire hand posture [36], interactions are performed by the positions of the two fingers, hence fewer computation is ensured.

Using the libraries of OpenCV and OpenGL, the technique is implemented in a project; IGF. Each time, the dynamic positions of the fingers in a scanned image frame are traced using an ordinary camera. By the relative 2D positions of the fingers, a 3D interaction is performed inside the designed VE. In a normal lighting environment, the system is evaluated for a total of 570 interactions by 15 participants. The satisfactory accuracy rate (94.3%) approves suitability of the technique in VR applications.

The paper is organized into 8 main sections. Section II is about the previous related research. The technique with its algorithm and mathematical details is covered in Section III. Details about the interactions are covered in Section IV. Implementation with evaluation details is explained in Section V. A comparative analysis is performed in Section VI whereas applicability in games is investigated in Section VII. The last section of the paper is about conclusion and future work.

\* Corresponding author.

E-mail address: visitrais@yahoo.com

## II. LITERATURE REVIEW

With the privilege of interaction, a VR-user maintains the belief of being there and acknowledges a VE as real as it gets. To achieve a high degree of naturalness, the interface of a VR system should be simple and consistent [37]. The conventional way of interaction with keyboard and mouse are insufficient to properly engross users [11]. As human gestures are flexible and natural [38], therefore gesture/posture based interactions are suitable for VR applications [15]. Various gesture-based techniques have been proposed in the literature of VR for intuitive interactions. The magnetic tracker's based system [39] provides 6-DOF for selection and DeSelection necessitates the use of both hands simultaneously. The haptic feedback interface for virtual objects manipulation of [40] needs the cumbersome setup of wires. The system is useable only inside a restricted workspace. The approach of different gestures for different commands is followed in the system; IVEAS (*Immersive Virtual Environment Authoring System*) designed by Lee et al. [41]. The technique is applicable for the basic interactions, however, the use of CyberGlove and Polhemus Fastrak for the detection of the finger joints and the required setup adds in discomfort. The tangible 3D system designed by Kim et al. [42] for the manipulation of 3D objects supports only five static gestures of hand. Reifinger et al. [43] presented their infrared based tracking system where infrared markers are used to track both static and dynamic hand gestures. The system is applicable in a wide lighting range but needs an array of six IR (Infrared) cameras besides a high-cost Head Mounted Display (HMD). Valuable research has been investigated about the use of Fiducial markers in VR and AR systems [44]. Gestures of the hand are traced by the FingARtips system of Buchmann et al. [45] for the manipulation of a virtual object. The unique Fiducial markers are used to distinguish different fingers. The use of full body tracking device; *Microsoft Kinect* has also been investigated for 3D interactions [46, 47]. Although Kinect is applicable for tracking body-gestures [48], it cannot differentiate individual fingers [49]. The technique of Jin et al. [50] detects the fingers movements with the help of the multi-sensors Leap Motion Controller (LMC). The system supports faster navigation with a satisfactory accuracy rate if used with HMD. However, due to the limited workspace of LMC [51], the accuracy of the system falls if gestures are posed outside a short range. Summarizing the challenges of gesture-based interactions, Benko [52] pointed out the key issue of false tracking [41]. This research work is an attempt to minimize false tracking by recognizing gestures of the two fingers with the help of an ordinary camera.

## III. GIFT: THE PROPOSED TECHNIQUE

This research work intends to present a gesture-based direct interaction [11] technique for interactions in a VE. All the interactions are performed on the basis of the dynamic positions of the two fingers.

Instead of following the computationally costlier methods of gesture recognition, the two finger-caps of colour green (for index) and pink (for thumb) are used. Lest a same-colour background object is detected, first a Region Of Interest (*ROI\_Img*) is extracted from dynamically scanned Frame-Image (*Fr\_Img*) on the basis of skin color. The *ROI\_Img* is then thresholded for green color to trace tip of index finger. The *Hue*, *Saturation* and *Values* (HSV) color space are used for the detection and tracking of the finger-caps. The *VH* representing the user's position in the digital world is activated at the detection of the finger-caps. From the 2D position of the fingertip-thimbles, the Central Point (*CP*) and the central areas for Index and Thumb ( $CA_{index}$  and  $CA_{Thumb}$ ) are computed from the first five dynamic images. On the basis of the *CP*,  $CA_{index}$  and the  $CA_{Thumb}$ , the *Selection-Manipulation* (SM) zone is defined. The *SM* is a limited area around a user's hand within which gestures for *selection*, *rotation*, *translation* and *scaling* can be posed with ease, see Fig. 1(a).

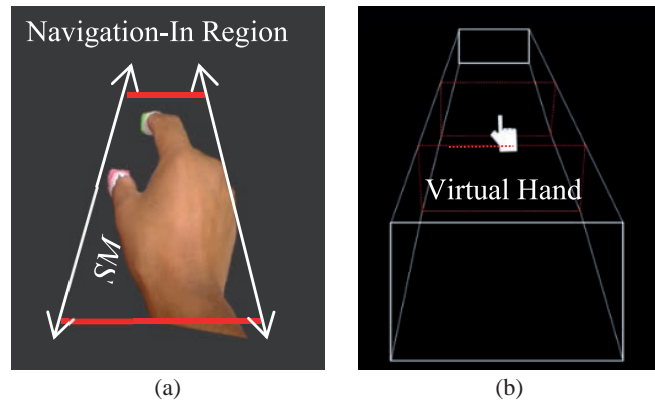


Fig. 1. (a) SM zone with Navigation regions (b) Perspective viewport of the VE.

The system works in three phases; *Exploration Phase* (EP), *Translation Phase* (TP) and *Manipulation Phase* (MP). Initially, the system starts with the EP and supports navigation and panning to explore a synthetic world. Translation and manipulation (scaling and rotation) are performed along an arbitrary axis in the TP and MP respectively. Switching between any two phases is activated by posing a distinct gesture. The three intuitive gestures of the two fingers; *Open-pinch*, *Closed-pinch* and *Cross-sign* are used to switch the system from one phase to another. Open-pinch is the posture where the two fingers are gently apart, see Fig. 2(a). With the posture of closed-pinch, tip of index touches the tip of index finger, see Fig. 2(b). Rotating the fingers along the look-at vector makes the cross-sign gesture, see Fig. 2(c).

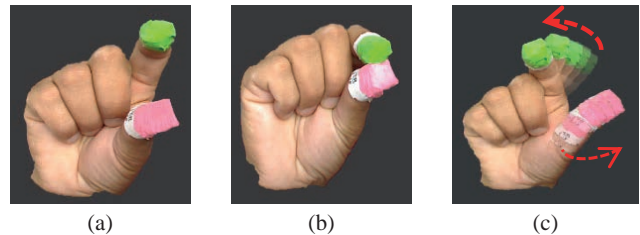


Fig. 2. The postures of (a) Open-pinch (b) Closed-pinch and (c) Cross-sign.

The open-pinch and closed-pinch gestures are used to activate MP and TP respectively. The cross-gesture is to switch the system back to the default EP. The object to be manipulated is inscribed inside two concentric cubes, as shown in Fig. 3. The outer cube inscribing the whole of an object specifies the Outer Aura (OA). The inner cube marking the central portion of an object represents the Inner Aura (IA). To translate an object, the IA selection of the object needs to be performed using the closed-pinch gesture. Similarly to activate manipulation of the object, the object should be selected for the OA using the open-pinch gesture. In order to visualize the switching between different phases, the appropriate cube around an object is highlighted.

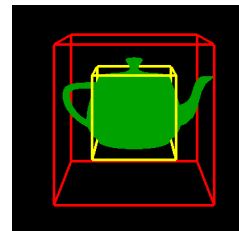


Fig. 3. Object inscribed inside two concentric cubes.

To explore a synthetic world, the system starts with the EP and supports only navigation and panning. Navigation and panning are performed outside the SM zone. Dynamic Areas (DAs) of the fingers are

compared with CAs for going inside into the VE (forward navigation) or for reverse navigation (backward navigation). Within the SM zone, translation and manipulation are performed based on the attained state of the system (TP or MP).

### A. The Architecture of the Proposed Technique

Besides rendering a virtual scene at the front-end, the system extracts the *ROI\_Img* based at the back-end. The *ROI\_Img* is converted to HSV and is then thresholded for a broader range of green colour. At the detection of the thimble of the index finger, the *ROI\_Img* is split horizontally at the center of the cap of the index finger to get the Down-Section Image (*DS\_Img*), see Fig. 4(b).

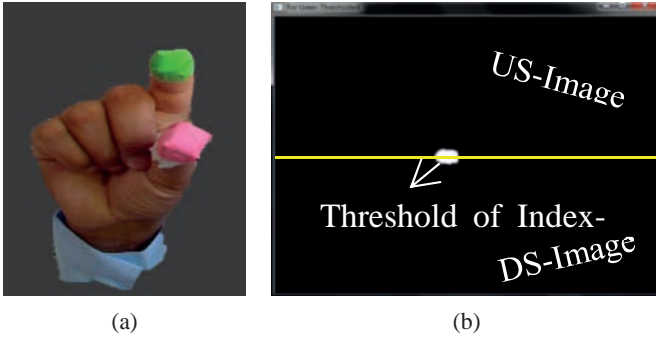


Fig. 4. (a) A hand pose in the *Fr\_Img* and (b) partition of the image into up and down sections with threshold of Index.

The *DS\_Img* is thresholded for pink color to find out the thumb-tip in the image. When both the fingers are traced, coordinates mapping between the fingers in image frame and the VH in the VE is initiated. As long as both the caps are visible, the VH can move freely with the movements of the fingers to access a 3D object. Navigation is performed by tracing a change in areas of the finger-caps while panning is performed on the basis of dynamic positions of the fingers.

If an object is selected with the open-pinch posture (OA selection), the MP is activated. Similarly, bringing the VH over an object and posing a closed-pinch (IA selection), the object is selected for translation (TP). User is informed about either of the states by highlighting the respective cube inscribing the object. Schematic of the entire system is shown in Fig. 5.

### B. Image Segmentation

Image segmentation refers to the extraction of a set of pixels [53]. At the outset, *ROI\_Img* is extracted to avoid the chances of false detection if similar colors are present in the background. The *ROI\_Img* is supposed to be the most probable section of *FR\_Img* containing both of the fingers. As the *YCbCr* space is best to differentiate between skin and non-skin colors [54], therefore the *YCbCr* model is followed for skin color segmentation.

$$Fr\_Img\_YCbCr \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.2 \left( \frac{219}{255} \right) & 0.7 \left( \frac{219}{255} \right) & 0.08 \left( \frac{219}{255} \right) \\ -\left( \frac{0.212}{1.8} \right) \left( \frac{224}{255} \right) & \left( \frac{224}{255} \right) - \left( \frac{0.7}{1.8} \right) \left( \frac{224}{255} \right) & 0.5 \left( \frac{224}{255} \right) \\ 0.5 \left( \frac{224}{255} \right) & -\left( \frac{0.7}{1.5} \right) \left( \frac{224}{255} \right) & -0.6 \left( \frac{224}{255} \right) \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

After getting the binary image of the *FR\_Img*, see Fig. 6, the *ROI\_Img* with rows 'm' and columns 'n' is extracted from the *FR\_Img* using our designed algorithm [55] as,

$$ROI\_Img(m, n) = \left( \begin{array}{c} \bigcup_{r=D_m}^{Fr\_Img.Row(0)} (Fr\_Img) \\ \bigcup_{c=L_m}^{Fr\_Img.Column(R_m)} (Fr\_Img) \end{array} \right) \quad (2)$$

where  $L_m$ ,  $R_m$  and  $D_m$  represents *Left-most*, *Right-most* and *Down-most* skin pixels.

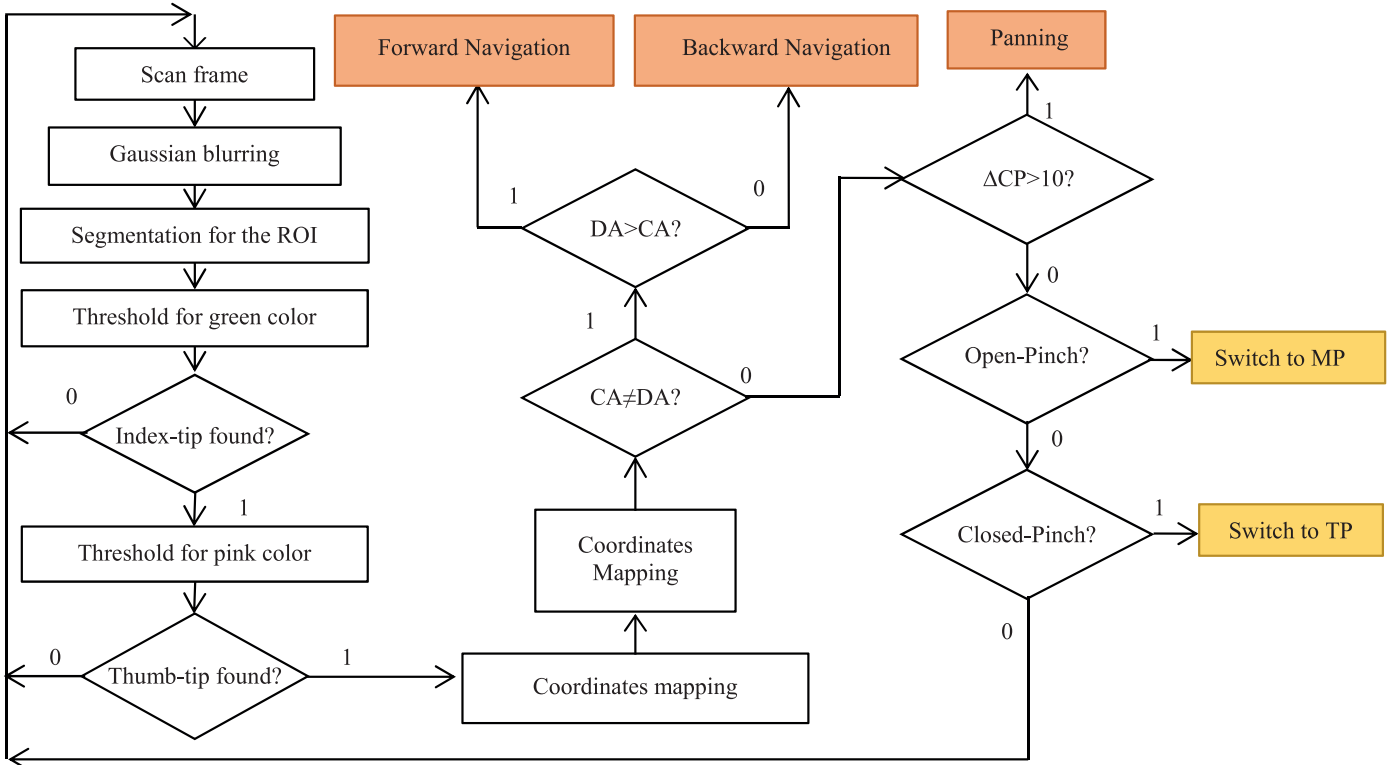


Fig. 5. Schematic of the proposed system.

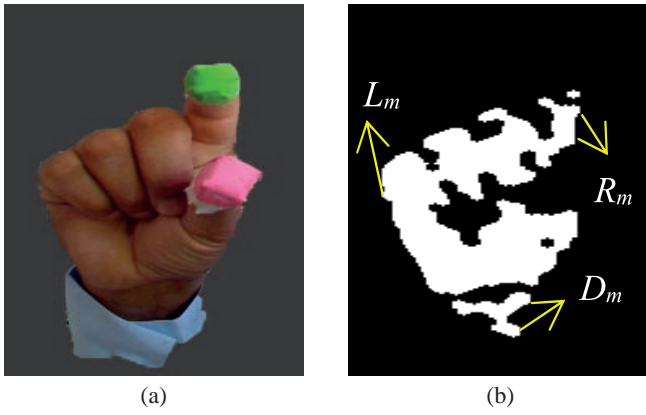


Fig. 6. The  $Fr\_Img$  in (a) RGB and (b) after conversion into YCbCr.

The segmented  $ROI\_Img$  is then thresholded for green and pink colors using the HSV color space.

$$ROI\_Img = \begin{cases} 1 & \text{if } 47 \leq ROI\_Img.H(x,y) \leq 94 \\ 1 & \text{if } 100 \leq ROI\_Img.S(x,y) \leq 187 \\ 1 & \text{if } 102 \leq ROI\_Img.V(x,y) \leq 255 \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

### C. Coordinates Mapping

A challenging task in the research was to harmonize the pixels representing the finger-caps with the coordinates of the VH. The coordinate system of OpenCV is different from the coordinate system of OpenGL. In OpenCV, a frame image starts with  $O(0,0)$  at top left. The origin  $O(0,0,0)$  of the OpenGL lies at the centre of a rendered frame, see Fig. 7.

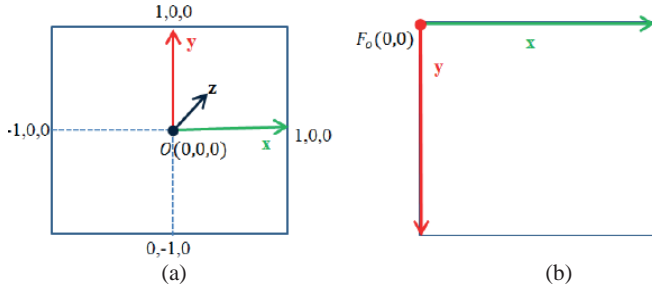


Fig. 7. The coordinates of (a) OpenGL and (b) OpenCV.

To harmonize the dissimilar coordinate systems, we devised our own mapping function,  $f$  [55]. The image frame is virtually split into four regions  $R_1$  to  $R_4$  as shown in Fig. 8. Mapping is made by the corresponding function taking 'x' and 'y' of a pixel of  $R_n$  as independent variables. The OpenCV pointer variable; CP is used to locate and move the virtual hand based on the position of both the fingers.

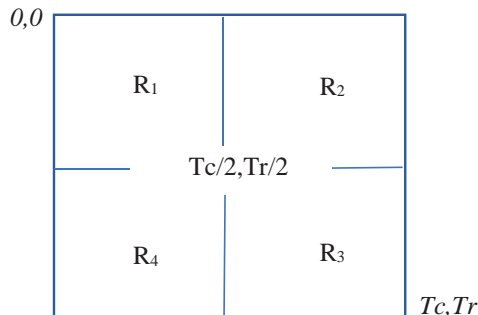


Fig. 8. The virtual division of an image frame.

If  $CP_i \in \mathbb{R}^2$  represents the CP in the initial frames and  $CP_d \in \mathbb{R}^2$  represents the dynamic position of the CP in any scanned frames, then the virtual hand's position;  $VHP: VHP \in \mathbb{R}^3$  is calculated as,

$$CP.x = (Index\_Pos.x + Thumb\_Pos.x)/2 \quad (4)$$

$$CP.y = (Index\_Pos.y + Thumb\_Pos.y)/2 \quad (5)$$

$$\Delta Px = (CPd.x - CPi.x) \quad (6)$$

$$\Delta Py = (CPd.y - CPi.y) \quad (7)$$

$$w(x; y) = ((CPx/Tc)k, (Cpy/Tr)k) \quad (8)$$

$$VHP(x, y) = w(CPf, CPd) \quad (9)$$

Where  $k$  is the speed constant; greater the value of  $k$  speedier will be the movements of the VH. The value of 'Tc' and 'Tr' represents the total number of columns and rows respectively. We kept a moderate value of  $k$  in the IGF project. Furthermore, to ignore the unintentional movements of the fingers the mapping is performed only when the value of  $\Delta Px$  or  $\Delta Py$  is greater than five pixels. The mapping process is shown in Fig. 9.

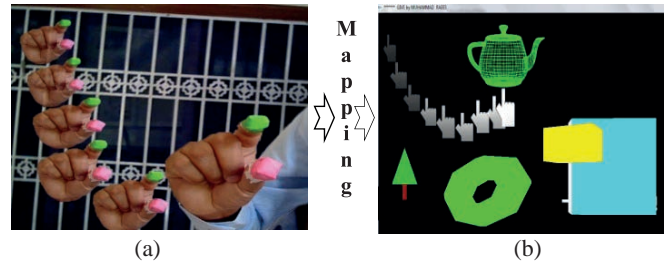


Fig. 9. The mapping between (a) the fingers move in image frame to (b) virtual hand in VE.

## IV. INTERACTIONS SUPPORT

All the basic 3D interactions are performed by the intuitive gestures of the fingers. The 2D positions of the fingers traced in the initial frames are supposed to be in a moderate range from the camera. Hence, the zone; SM is set as soon as the fingers' tips are recognized by the system.

### A. Navigation

Navigation is to explore a VE. Mostly, a user needs to navigate to a proper location before performing selection or manipulation. In the proposed technique, forward or backward movement of the hand outside the SM zone performs navigation. As plausible, the forward navigation is carried out by the forward movement of the hand (see Fig. 10). Similarly, backward navigation is performed by the backward movement of hand outside the SM zone.

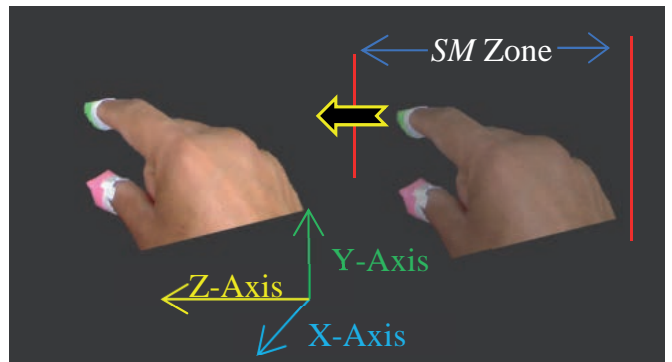


Fig. 10. The forward movement of hand for forward-navigation.

To deduce the forward or backward hand movements in a scanned 2D image, the DAs (Dynamic Areas) of the fingers-caps are calculated and are compared with the CAs at runtime. By crossing the forward limit of the SM in the z-axis, DAs increase, see Fig. 11(b).

$$DA_x = (y)(CA_x) \quad (10)$$

where,  $x = \{\text{index, thumb}\}$  and  $y > 1$ .

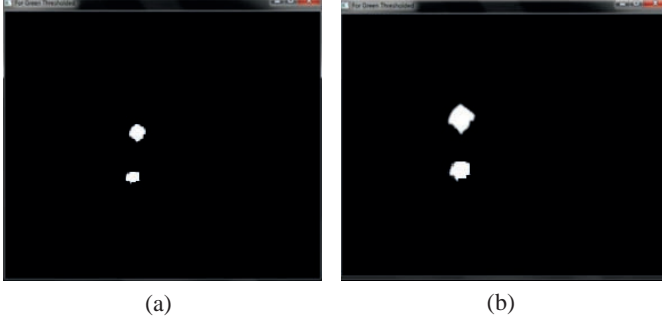


Fig. 11. (a) The CAs in initial image frame and (b) the DAs after forward hand movement.

To avoid the possibility of unintended increase/decrease in the DAs, the Fingers Average Dynamic (FAD) positions along the x-axis (FAD.x) and y-axis (FAD.y) are also checked against CP.x and CP.y, as clear from the following pseudo-code.

```

if (DAIndex > CAIndex AND DAThumb > CAThumb) AND
    (FAD.x ≤ CP.x + 10 AND FAD.x ≥ CP.x - 10)
    AND (FAD.y ≤ CP.y + 10 AND FAD.y
        ≥ CP.y - 10)
    Forward Navigation
if (DAIndex < CAIndex AND DAThumb < CAThumb) AND
    (FAD.x ≤ CP.x + 10 AND FAD.x ≥ CP.x - 10)
    AND (FAD.y ≤ CP.y + 10 AND FAD.y ≥ CP.y - 10)
    Backward Navigation

```

### B. Selection DeSelection

Selection is choosing an object for interaction. Selection for scaling and rotation is made by bringing the VH over the object and posing an open-pinch gesture. With this, the OA selection is performed whereas the outer cube inscribing the object is highlighted. Similarly, an object is selected for translation by posing a closed-pinch to perform the IA selection. By posing a cross gesture by the two fingers deselects the object. The system recognizes the gesture when the fingers exceed a specified threshold along x-axis from the horizontal central positions of the fingers. If *ID* and *TD* represent Index-Dynamic and Thumb-Dynamic positions then, pseudo code for the DeSelection is given as,

```

If (ID.x > CP.x + 30 AND TD.x + 30 < CP.x)
    AND (ABS(ID.y - CP.y) > 10 AND (ABS(TD.y - CP.y) > 10)
    AND ABS(DAIndex - CAIndex) < 20
    AND ABS(DAThumb - CAThumb) < 20
    DeSelect

```

Once an object is deselected, the system switches back to the default EP.

### C. Translation

Translation is changing the position of a virtual object along x, y and/or z-axis in a VE. Like gripping an object in the real world, the closed-pinch gesture is used to select an object for translation. To perform translation of an object, the system needs to be in the TP. The transition from the EP to the TP is made by hovering the VH over an

object and posing the pinch-gesture. Besides switching from the EP to TP, the IA selection of the object is performed. After the selection, translation is performed in the VE by the movements of the hand.

The following condition is checked to ensure the closed-pinch posture.

```

if (ABS(DAindex - CAindex) < 20
    AND (ABS(DAThumb - CAThumb) < 20
    AND (ABS(ID.x - TD.x) < 10)
    AND (ABS(ID.y - TD.y) < 10)
    Select for Translation

```

Unless the fingers' tips are made apart, the IA selection of the object remains intact. During translations, the dynamic coordinates of the VH are assigned to the selected object to translate the object by the free movement of hand accordingly. After the IA selection of an object *Obj*, the 3D coordinates of the *Obj* are obtained as,

$$\begin{pmatrix} Obj.x \\ Obj.y \\ Obj.z \end{pmatrix} = \begin{pmatrix} VHP.x \\ VHP.y \\ VHP.z \end{pmatrix} \quad (11)$$

### D. Rotation

Within the SM zone, a selected object is rotated using the ChordBall technique [56]. As perceivable, moving the VH along the x-axis with the open-pinch rotates the object along the y-axis. Similarly, movement of the VH along y-axis rotates the object along the x-axis. The angle  $\theta$  of rotation is calculated using the distance ( $d$ ) of the fingers from the central position (*CP*). Greater the distance, larger will be the angle of rotation.

$$d = \sqrt{(CP.x - FAD.x)^2 + (CP.y - FAD.y)^2} \quad (12)$$

$$\theta = d / (Tc/2) \quad (13)$$

The following algorithm is followed to perform rotation about an axis,

```

if (ABS(FAD.y - CP.y) < 10)
    AND (ABS(DAindex - CAindex) < 20)
    AND (ABS(FAD.x - CP.x) > 20)
    Rotation about y-axis
if (ABS(FAD.x - CP.x) < 10)
    AND (ABS(DAindex - CAindex) < 10)
    AND (ABS(DAThumb - CAThumb) < 20)
    AND (ABS(FAD.y - CP.y) > 20)
    Rotation about x-axis
if (ABS(FAD.x - CP.x) > 20)
    AND (ABS(DAindex - CAindex) > 20)
    AND (ABS(DAThumb - CAThumb) < 20)
    AND (ABS(FAD.y - CP.y) > 20)
    Rotation about z-axis

```

### E. Scaling

After the OA selecting with the open-pinch, an object can be scaled up or down. With the activation of the MP, the initial distance between the two fingers is traced. In the MP, a selected object is scaled or scale-down by comparing the dynamic distances with the initial distance between the fingers. A gentle increase in the distance between the thumb and index fingers scales up the object and vice versa. If the gesture is parallel to the x-axis then up or downscaling along the x-axis is performed. Scaling or downscaling along the y-axis is performed if the gesture is parallel to the y-axis as shown in Fig. 12(a) and Fig. 12(b).

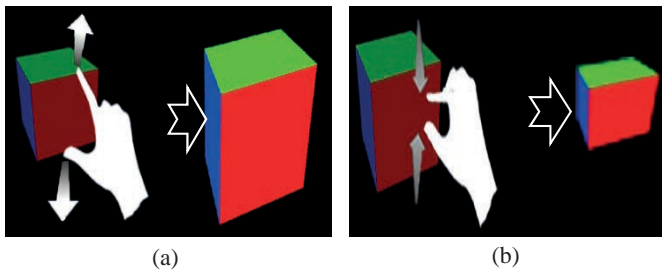


Fig. 12. (a) Gesture for scaling along y-axis and (b) Gesture for downscaling along y-axis.

Scaling or scale-down about the z-axis is enacted by increasing or decreasing the diagonal distance respectively, see Fig. 13.

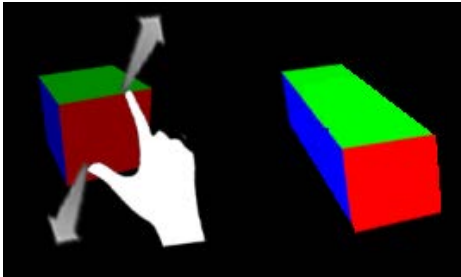


Fig. 13. The scaling along the z-axis.

## V. IMPLEMENTATION AND EVALUATION OF THE TECHNIQUE

To systematically evaluate the proposed approach, the technique is implemented in the case-study application; IGF (Interaction by the Gestures of Fingers). The two open libraries; OpenGL and OpenCV are used for the front-end rendering and for the back-end image processing respectively. A Corei3 laptop with 2.30 GHz processor and 4GB RAM was used for the implementation and evaluation. The built-in camera of the laptop (resolution 640x480) was used to capture the image streams at runtime. Fifteen participants, all male, of ages between 25 and 40 (mean=31, SD=6) performed the six predefined tasks. Each participant performed two trials of the tasks. Before starting the evaluation session, participants of the evaluation were introduced to the system and pre-trials were performed by the users.

In the IGF application, the VH moves freely in the environment where the coordinates of virtual camera change with the fingers movement accordingly. The z-axis of the VH is intentionally kept constant so that to be visible for interaction everywhere in the VE. Activation of the entire system is conditioned to the visibility of the fingers' caps. The users are constantly informed by the text "DETECTED" displayed in the upper centre part of the scene. Similarly, a user is notified both with text and a beep-signal (audio) when an interaction is activated, see Fig. 14.

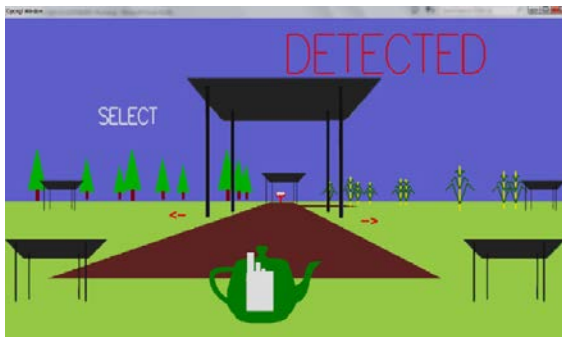


Fig. 14. The virtual scene for evaluation of the technique.

## A. Testing Environment and the Experimental Tasks

The 3D environment designed for the evaluation of the technique contains six tables; two on either side and two in the middle. The VH made of cubes represents the user's position in the environment. For easy noticing, endpoint of the scene is marked by a board with the text "Stop". To enhance immersivity, the scene is designed with different 3D objects at different positions. By hitting the *Enter* key, the system restarts with a new trial/task.

Participants were asked to perform the following six tasks in the designed 3D environment. The tasks cover the basic interaction operations, i.e. *Selection*, *DeSelection*, *Translation*, *Scaling*, *Rotation* and *Navigation*. A 3D object (Teapot) is rendered in the mid of the scene for selection and manipulation.

**Task-1:** Select the object, translate it to the far left table and then deselect it.

**Task-2:** Select the object, translate it to the far right table and then deselect it.

**Task-4:** Navigate to the endpoint marked by the stop-board and then navigate back to the starting point.

**Task-5:** Select the object, scale it up/down along x-axis/y-axis and then deselect it.

**Task-3:** Select the object, scale it up/down along the z-axis and then deselect it.

**Task-6:** Select the object, rotate it along the x-axis, y-axis and z-axis and then deselect it.

While performing these tasks, selection and deselection are evaluated five times, rotation three times while translation, navigation and scaling two times each in a single trial. Missed or false detection of the system, after posing the specified gestures, were counted as errors. With this setup, overall accuracy rate for all the interaction tasks, as shown in Table I, is 94.3%. Comparatively more errors occurred during translation and navigation which were due to quicker move of hand or both of the fingers at the same time. In such cases, the camera misses tips of the fingers. This implies that the system performance can be improved with a high-efficiency camera.

TABLE I. STATISTICS OF THE INTERACTION TASKS

Interaction	Correct	Total	%age
Selection	146	150	97.3
DeSelection	57	60	95
Translation	55	60	91.6
Scaling	86	90	95.5
Rotation	56	60	93.3
Navigation	140	150	93.3
<b>Total</b>	<b>540</b>	<b>570</b>	<b>94.3</b>

## B. Learning Effect

The learning effect was assessed from the errors occurrence rate. The paired two-sample T-test was used to analyze differences in means of the two trails. With the null hypothesis ( $H_0$ ), we assumed that the mean difference ( $\mu_d$ ) is 0. The hypothesis was rejected as there was a significant difference between the outcomes of Trail-1 (M=92.8, SD=1.7) and Trail-2 (M=96, SD=1.29) conditions; ( $t(5)=-3.23$ ,  $p=0.0091$ ). The graph shown in Fig. 15 and Fig. 16 indicate the vivid decrease in errors.

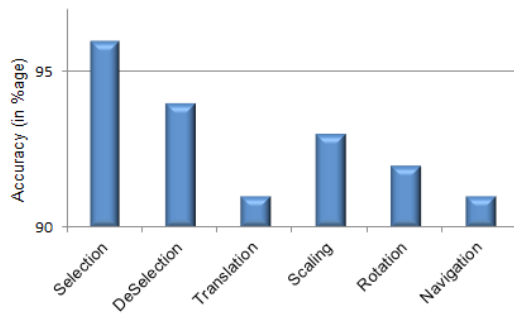


Fig. 15. Accuracy of the interactions performed in trial-1.

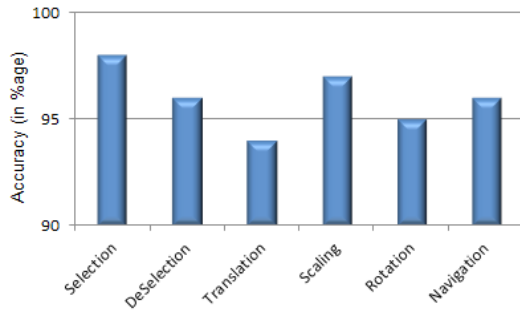


Fig. 16. Accuracy of the interactions performed in trial-2.

### C. Subjective Analysis

To analyze the response of the participants, a questionnaire was presented to the users at the end of the evaluation session. The questionnaire was to measure the three factors; *Ease of use*, *Fatigue* and *Suitability* of the technique in VEs. The post-assessment questionnaire is shown in Table II.

TABLE II. THE POST-ASSESSMENT QUESTIONNAIRE ABOUT THE FOUR FACTORS

Question No.	Statement	Your response (tick one)?				
		Strongly agree	Agree	Indifferent	Disagree	Strongly disagree
1	As a whole, the system was easy to use.					
2	No fatigue, stress or strain was felt during or after performing the interaction tasks					
3	The technique is suitable for VR/AR applications.					

The percentage of users' response to the three factors is shown in Fig. 17.

## VI. COMPARATIVE STUDY

Following the interaction standards presented by various researchers [57-59], the proposed approach is systematically compared with the recent state-of-the-art interaction techniques. Based on the available information (acquired from the literature), a score ('1') is assigned to a technique if a particular interaction rule ( $R_i$ ) is followed. The final score is calculated as,

$$Score = \sum_{i=1}^N R_i / N \quad (14)$$

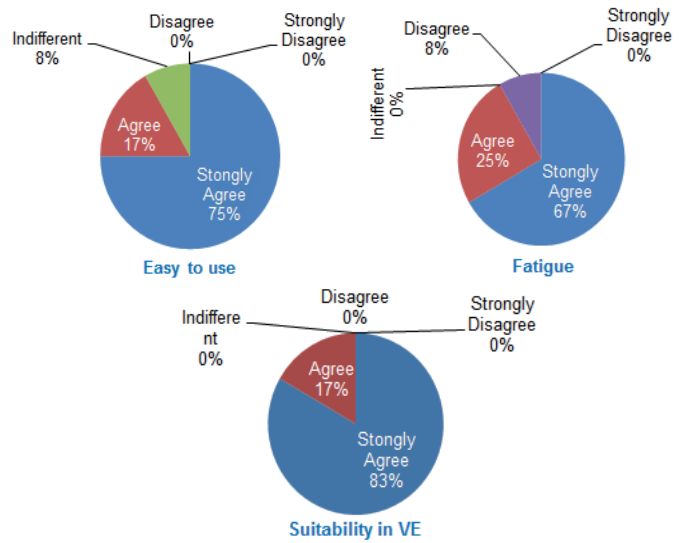


Fig. 17. Response of the participants (in %) about the three factors.

Where  $N$  is the total number of the standard rules of 3D interactions (see Table III).

TABLE III. THE BASIC FIVE STANDARDS OF INTERACTIONS

Standards	Description
$R_1$	Interactive support
$R_2$	Reliable sensory feedback
$R_3$	Simple and inexpensive
$R_4$	Useable while sitting
$R_5$	Wireless connection

A total of eight state-of-the-art interaction techniques (including the proposed approach) were selected for comparative analysis. Details of the techniques with the *possible challenge(s)* are presented in Table IV. After extracting information from the research works about the five standards, the final score was computed, see Table V.

TABLE V. EVALUATION OF THE TECHNIQUES BASED ON THE FIVE RULES

Technique No.	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	Score
1	1	1	0	1	0	0.6
2	1	1	0	1	0	0.6
3	1	1	0	1	0	0.6
4	1	1	0	1	0	0.6
5	1	1	0	1	0	0.6
6	1	1	1	1	0	0.8
7	1	1	0	1	0	0.6
8	1	1	1	1	1	1

## VII. APPLICABILITY IN GAMES

The game technology is shifting from 2D graphics to realistic 3D VR. Recent research works have proved that gesture-based games are suitable for games [67] [68]. The free hand and/or fingers gestures have promising potential not only in education, training and medical application [69-71] but also in the contemporary VR games [67]. It has been proved that the use of gestures improves overall engagement of the gamers [72] [73]. To make VR games interesting and intuitive, a number of gesture-based techniques have been proposed. The technique proposed by [67] and [68] are based on Leap motion and Nintendo Wii controllers. However such systems suffer from the limited work space of the controllers and training of gestures [68].

The proposed technique is applicable in the games where the gamer

TABLE IV. THE DETAILS OF THE CONTEMPORARY INTERACTION TECHNIQUES

Tech. No.	Author(s)	Year	Tracking device used	Title	Possible challenge(s)
1	Linn, Andreas [60]	2017	HTC Vive with Tobii Eye-tracker	Gaze Teleportation in Virtual Reality	Highly Sensitive to calibration errors
2	Thammathip Piumsomboon [61]	2017	HMD with eye-tracker	Exploring Natural Eye-Gaze-Based Interaction for Immersive Virtual Reality	Support only selection
3	Mohamed Khamis et al. [62]	2018	HTC Vive eye-tracker	VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements	Selection while walking is erroneous and time-consuming
4	Stellmach et al. [63]	2012	Tobii T60 eye-tracker	Designing Gaze-based User Interfaces for Steering in Virtual Environments	Continuous gazing at destination point may lead to asthenopia
5	Wen-jun Hou et al. [64]	2018	HTC Vive	User Defined Eye Movement-Based Interaction for Virtual Reality	Multiple tasks by the same gestures may lead to fuzzy input
6	G. Prabhakar and P. Biswas [65]	2018	Eye-gaze trackers	Eye Gaze Controlled Projected Display in Automotive and Military Aviation Environments	Supports only pointing and selection
7	Qi Sun et al. [66]	2018	HMD with SMI gaze tracker	Towards Virtual Reality In_nite Walking: Dynamic Saccadic Redirection	Suffers from tracking latency and redirection during blinking.
8	M. Raees and S. Ullah	2018	Ordinary camera	The proposed technique	Lighting is required to trace the finger-tips

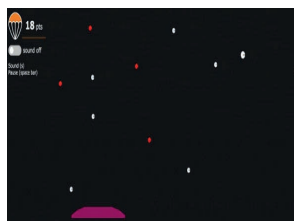
needs to select and/or controls actions of an avatar/character in the game environment. Unlike other gesture-based systems [74], the proposed technique neither keeps feature sets [75-77] nor needs searching and classification [78-80], hence ensure speed in playing games. To properly assess applicability of the proposed technique in the contemporary VR games we selected the three well-liked games; see Fig. 18.



(a)



(b)



(c)

Fig. 18. The scenes of (a) Alien invasion (b) Play breads puzzle and (c) Skyfall games.

Skyfall is a simple game created using the physics engine; planck.js [81]. There are three types of balls in the game that fall from the top of the screen in a random order. White and green balls have worth +10 points while red balls are of -10 points. Players need to earn more

points by moving a paddle to catch the good balls (white and green balls). However, at the same time a gamer should avoid bad balls (red balls). The GIFT technique is applicable to control the paddle by the movements of the fingers.

Similarly, the Beads Puzzle [82] is a ball-shooter game where a gamer has to fire a ball at a time. Direction of the cannon is controlled by the mouse movements. A fire is made by the mouse down event. The game can be played with the proposed technique more easily than with a mouse.

Unity's Alien Invasion [83] is another popular game where a player has to defend a city from aliens. The player needs to control a character that is able to collect bombs and fire. With the proposed technique, a gamer can more efficiently control actions of the character. One suitable configuration of the game with the technique is 1) to control movements of the character with the dynamic positions of the fingers. 2) To use the *Pinch-gesture* for collecting the bombs and 3) to use the *Cross-gesture* for firing the bullets. An additional research is required to assess applicability of the technique in HMD based VR games.

## VIII. CONCLUSION AND FUTURE WORK

To cope with the pace of VR developments in various fields, simple and natural interfaces are in need. With this contribution, we proposed a novel gesture-based interaction technique which needs no expensive device other than an ordinary camera and pieces of paper. Simple and intuitive gestures are used to ensure naturalism while performing the basic interaction; navigation, selection, scaling, rotation and translation. Experimental results show that the proposed approach has reliable recognition and accuracy rates. The system neither needs training of images nor use any feature extraction, hence guarantying fast preprocessing. The proposed system is feasible and well suitable



in a wide spectrum of HCI including virtual prototyping, 3D gaming, robotics, industrial architecture and simulation. The work also covers the smooth integration of image processing and VE and can be used in the designing of a realistic VR application.

We believe that the technique is suitable for the basic 3D interactions, however the technique suffers from illumination variation [15]. Moreover, the technique is not applicable for the two-player shooter games. An additional research is required to reduce dependency of the system on the lighting condition and to make the technique suitable for the collaborative VE. In future, we are determined to enhance the technique for the emerging augmented and mixed VR setups.

## REFERENCES

- [1] Uddin, M. Taufeeq, and M. A. Uddiny, "Human activity recognition from wearable sensors using extremely randomized trees." In International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-6, (2015).
- [2] J. Ahmad, N. Sarif, J. T. Kim, and T.S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home." *Indoor and built environment*, 22, no. 1, pp. 271-279, (2013).
- [3] Zhan, Yi, and T. Kuroda, "Wearable sensor-based human activity recognition from environmental background sounds," *Journal of Ambient Intelligence and Humanized Computing* 5, no. 1, pp. 77-89, (2014).
- [4] J. Ahmad, "Security architecture for third generation (3g) using gmhs cellular network," In International Conference on Emerging Technologies, pp. 74-79, (2007).
- [5] J. Ahmad, and M. A. Zeb, "Security enhancement for e-learning portal," *International Journal of Computer Science and Network Security*, 8, no. 3, pp. 41-45, (2008).
- [6] J. Ahmad and Y. A. Rasheed, "Collaboration achievement along with performance maintenance in video streaming," In Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria, vol. 2628, p. 18, (2007).
- [7] J. Ahmad, and A. Shahzad, "Multiple facial feature detection using vertex-modeling structure," In Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria, vol. 2628, (2007).
- [8] J. Ahmad, S. Kim, and B. J. Yun. "Assembled algorithm in the real-time H. 263 codec for advanced performance." In Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, pp. 295-298, (2005).
- [9] J. Ahmad, and S. Kim, "Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication," *ICGST International Journal on Graphics, Vision and Image Processing* 5, no. 9, pp. 27-32, (2005).
- [10] J. Ahmad and S. Kim, "Algorithmic Implementation and Efficiency Maintenance of Real-Time Environment using Low-Bitrate Wireless Communication," In Proc. of the 4th IEEE Workshop on Software Technologies for Future Embedded and Ubiquitous Systems, and the Second International Workshop on Collaborative Computing, Integration, and Assurance (SEUS-WCCIA'06), pp. 81-88, (2006).
- [11] S. S. Rautaray, "Real Time Hand Gesture Recognition System for Dynamic Applications," *Int. J. UbiComp*, vol. 3, no. 1, pp. 21-31, (2012).
- [12] Sahane, M. S., H. D. Salve, N. D. Dhawade, and S. A. Bajpai, "Visual Interpretation Of Hand Gestures For Human Computer Interaction," *environments (VEs)*, 2, p.53, (2014).
- [13] R. Itkarkar and A. Nandy, "A Study of Vision Based Hand Gesture Recognition for Human Machine Interaction," vol. 1, no. 12, pp. 48-52, (2014).
- [14] Pirkker, Johanna, M. Pojer, A. Holzinger and C. Gütl, "Gesture-based interactions in video games with the leap motion controller," In International Conference on Human-Computer Interaction, pp. 620-633, (2017).
- [15] Kim, Kwangtaek, J. Kim, J. Choi, J. Kim and S. Lee, "Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions," *Sensors* 15, no. 1, pp. 1022-1046, (2015).
- [16] Koller, Dieter, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, "Real-time vision-based camera tracking for augmented reality applications," In Proceedings of the ACM symposium on Virtual reality software and technology, pp. 87-94, (1997).
- [17] J. Ahmad, Y. Kim, and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," In International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6, (2014).
- [18] J. Ahmad, J. T. Kim, and T-S Kim, "Development of a life logging system via depth imaging-based human activity recognition for smart homes," In Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea, vol. 19, (2012).
- [19] J. Ahmad, M. Z. Uddin, and T-S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Transactions on Consumer Electronics* 58, no. 3, pp. 863-871, (2012).
- [20] Kamal, Shaharyar, J. Ahmad and D. Kim, "Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM," *J. Electr. Eng. Technol* 11, no. 3, pp. 1921-1926, (2016).
- [21] J. Ahmad, S. Lee, J. T. Kim and T-S. Kim, "Human activity recognition via the features of labeled depth body parts," In International Conference on Smart Homes and Health Telematics, pp. 246-249, (2012).
- [22] J. Ahmad, and S. Kamal, "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services," In International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 74-80, (2014).
- [23] Huang, Qiao, J. Yang, and Y. Qiao, "Person re-identification across multi-camera system based on local descriptors," In Sixth International Conference on Distributed Smart Cameras (ICDSC), pp. 1-6, (2012).
- [24] J. Ahmad and S. Kim, "Global security using human face understanding under vision ubiquitous architecture system," *World academy of science, engineering, and technology* 13, pp. 7-11, (2006).
- [25] Farooq, Faisal, J. Ahmed and L. Zheng, "Facial expression recognition using hybrid features and self-organizing maps," In International Conference on Multimedia and Expo (ICME), pp. 409-414, (2017).
- [26] Farooq, Adnan, J. Ahmad and S. Kamal, "Dense RGB-D map-based human tracking and activity recognition using skin joints features and self-organizing map," *KSII Transactions on Internet and Information Systems (TIIS)* 9, no. 5, pp.1856-1869, (2015).
- [27] Kamal, Shaharyar, and J. Ahmad, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian Journal for science and engineering* 41, no. 3, pp.1043-1051, (2016).
- [28] J. Ahmad, S. Kamal and D. Kim, "A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems," *International Journal of Interactive Multimedia & Artificial Intelligence* 4, no. 4, (2017).
- [29] Yoshimoto, Hiromasa, N. Date and S. Yonemoto, "Vision-based real-time motion capture system using multiple cameras," In International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 247-251, (2003).
- [30] J. Ahmad, S. Kamal and D-S. Kim, "Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System," *KSII Transactions on Internet & Information Systems* 12, no. 3, (2018).
- [31] A. Shamaie and A. Sutherland, "Accurate recognition of large number of hand gestures," *Proc Iran. Conf. Mach. Vis. Image Process. Univ. Technol. Tehran*, (2003).
- [32] X. Yan and N. Aimaiti, "Gesture-based interaction and implication for the future," vol. 11, no. 1, p. 25, (2011).
- [33] K. Katsuragawa, A. Kamal, and E. Lank, "Effect of Motion-Gesture Recognizer Error Pattern on User Workload and Behavior," *Proc. 22nd Int. Conf. Intell. User Interfaces - IUI '17*, pp. 439-449, (2017).
- [34] P. H. An. "The challenge of hand gesture interaction in the Virtual Reality Environment: evaluation of in-air hand gesture using the Leap Motion Controller," <http://urn.fi/URN:NBN:fi:amk-2018091215089>, (Accessed on 4 Dec. 2018).
- [35] Yuan, Xiaobu, and J. Lu. "A Vision-Based Approach of Bare-Hand Interface Design in Virtual Assembly," In *Information Technology For Balanced Manufacturing Systems*, pp. 281-290, (2006).
- [36] P. Xu, "A Real-time Hand Gesture Recognition and Human-Computer

- Interaction System,” pp. 1–8, (2017).
- [37] C. Grönegress, “Designing Intuitive Interfaces for Virtual Environments,” *Simulation*, no. August, pp. 1–128, (2001).
- [38] K. S. Hale and K. M. Stanney, *Handbook of Virtual Environments: Design, Implementation, and Applications*. (2015).
- [39] Kiyokawa, Kiyoshi, H. Takemura, Y. Katayama, H. Iwasa, and N. Yokoya. “Vlego: A simple two-handed modeling environment based on toy blocks,” *Proc. Of ACM VRST*, 96, (1996).
- [40] J. Hua and H. Qin, “Haptic sculpting of volumetric implicit functions,” *Proc. - Pacific Conf. Comput. Graph. Appl.*, pp. 254–264, (2001).
- [41] Lee, Chan-Su, J. D. Choi, K. M. Oh, and C.J. Park. “Hand Interface for Immersive Virtual Environment Authoring System,” *In Proc. of the International Conference on Virtual Systems and MultiMedia*, pp. 361–366, (1999).
- [42] H. Kim, G. Albuquerque, S. Havemann, and D. D. W. Fellner, “Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling,” *Proc. 11th Eurographics Conf. Virtual Environ.*, pp. 191–199, (2005).
- [43] S. Reifinger, F. Wallhoff, M. Ablassmeier, T. Poitschke, and G. Rigoll, “Static and dynamic hand-gesture recognition for augmented reality applications,” *Human-Computer Interact. HCI Intell. Multimodal Interact. Environ. HCI 2007. Lect. Notes Comput. Sci.*, vol. 4552, pp. 728–737, (2007).
- [44] Rabbi, Ihsan, S. Ullah, and D. Khan. “Automatic generation of layered marker for long range augmented reality applications,” *Kuwait Journal of Science* 44, no. 3, (2017).
- [45] Buchmann, Volkert, S. Violich, M. Billinghurst, and A. Cockburn. “FingARtips: gesture based direct manipulation in Augmented Reality,” *In Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 212–221, (2004).
- [46] Song, Peng, W. B. Goh, W. Hutama, C. W. Fu, and X. Liu. “A handle bar metaphor for virtual object manipulation with mid-air interaction,” *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1297–1306, (2012).
- [47] S. Oprisescu and E. Barth, “3D hand gesture recognition using the hough transform,” *Adv. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 71–76, (2013).
- [48] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, “Analysis of the accuracy and robustness of the Leap Motion Controller,” *Sensors (Switzerland)*, vol. 13, no. 5, pp. 6380–6393, (2013).
- [49] M. Nabiyouni, B. Laha, and D. A. Bowman, “Poster: Designing effective travel techniques with bare-hand interaction,” *IEEE Symp. 3D User Interfaces 2014, 3DUI 2014 - Proc.*, no. March 2015, pp. 139–140, 2014.
- [50] H. Jin, Q. Chen, Z. Chen, Y. Hu, and J. Zhang, “Multi-LeapMotion sensor based demonstration for robotic refine tabletop object manipulation task,” *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 104–113, (2016).
- [51] Benko, Hrvoje. “Beyond flat surface computing: challenges of depth-aware and curved interfaces,” *In Proceedings of the 17th ACM international conference on Multimedia*, pp. 935–944, (2009).
- [52] A. V. Devadoss, R. Malaishamy, and M. Subramaniam, “Performance improvement using an automation system for recognition of multiple parametric features based on human footprint,” *Kuwait Journal of Science* 42, no. 1 (2015).
- [53] Phung, S. Lam, A. Bouzerdoum, and D. Chai “A novel skin color model in ycbcr color space and its application to human face detection,” *In Proceedings of International Conference on Image Processing*, vol. 1, pp. I-I, (2002).
- [54] M. Raees, S. Ullah, S. U. Rahman, and I. Rabbi “Image based recognition of Pakistan sign language,” *Journal of Engineering Research* , 4, no. 1 (2016).
- [55] M. Raees, S. Ullah, and S. U. Rahman. “VEN-3DVE: vision based egocentric navigation for 3D virtual environments,” *International Journal on Interactive Design and Manufacturing (IJIDeM)*, pp. 1–11, (2018).
- [56] M. Raees, S. Ullah, and S. U. Rahman. “CHORDBALL: A ROTATION TECHNIQUE FOR 3D VIRTUAL ENVIRONMENTS,” *Pakistan Journal of Science*, 69, no. 1, p. 85, (2017).
- [57] Alqahtani, A. Saeed, L. F. Daghestani, and L. F. Ibrahim. “Environments and System Types of Virtual Reality Technology in STEM: a Survey,” *International Journal of Advanced Computer Science and Applications (IJACSA)* 8, no. 6 (2017).
- [58] Zielasko, Daniel, S. Horn, S. Freitag, B. Weyers, and T. W. Kuhlen. “Evaluation of hands-free HMD-based navigation techniques for immersive data analysis.” *In 3D User Interfaces (3DUI)*, 2016 IEEE Symposium on, pp. 113–119, (2016).
- [59] Tapu, Ruxandra, B. Mocanu, and E. Tapu. “A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments.” *In Electronics and Telecommunications (ISETC)*, 2014 11th International Symposium on, pp. 1–4, (2014).
- [60] Linn, Andreas. “Gaze Teleportation in Virtual Reality,” (2017).
- [61] Piumsomboon, Thammathip, G. Lee, R. W. Lindeman, and M. Billinghurst. “Exploring natural eye-gaze-based interaction for immersive virtual reality,” *In Symposium on 3D User Interfaces (3DUI)*, pp. 36–39, (2017).
- [62] Khamis, Mohamed, C. Oechsner, F. Alt, and A. Bulling. “VRpursuits: interaction in virtual reality using smooth pursuit eye movements,” *In Proceedings of the International Conference on Advanced Visual Interfaces (AVI’18)*. ACM, New York, NY, USA, vol. 7. (2018).
- [63] Stellmach, Sophie, and R. Dachselt. “Designing gaze-based user interfaces for steering in virtual environments,” *In Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 131–138, (2012).
- [64] Hou, Wen-jun, K. Chen, H. Li, and H. Zhou. “User defined eye movement-based interaction for virtual reality,” *In International Conference on Cross-Cultural Design*, pp. 18–30, (2018).
- [65] Prabhakar, Gowdham, and P. Biswas. “Eye Gaze Controlled Projected Display in Automotive and Military Aviation Environments.” *Multimodal Technologies and Interaction*, 2, no. 1 (2018).
- [66] Sun, Qi, A. Patney, L.Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman “Towards virtual reality infinite walking: dynamic saccadic redirection,” *ACM Transactions on Graphics (TOG)* 37, no. 4, p.67, (2018).
- [67] Pirker, Johanna, M. Pojer, A. Holzinger and C. Gütl, “Gesture-based interactions in video games with the leap motion controller,” *In International Conference on Human-Computer Interaction*, pp. 620–633, (2017).
- [68] Kratz, Louis, M. Smith and F. J. Lee, “Wiizards: 3D gesture recognition for game play input,” *In Proceedings of the 2007 conference on Future Play*, pp. 209–212, (2007).
- [69] Khademi, Maryam, H. M. Hondori, A. McKenzie, L. Dodakian, C. V. Lopes and S. C. Cramer, “Free-hand interaction with leap motion controller for stroke rehabilitation,” *In CHI’14 Extended Abstracts on Human Factors in Computing Systems*, pp. 1663–1668, (2014).
- [70] Bizzotto, Nicola, A. Costanzo, L. Bizzotto, D. Regis, A. Sandri and B. Magnan, “Leap motion gesture control with OsiriX in the operating room to control imaging: first experiences during live surgery,” *Surgical innovation*, 1, no. 2, pp. 655–656, (2014).
- [71] Potter, L. Ellen, J. Araullo and L. Carter, “The leap motion controller: a view on sign language,” *In Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*, pp. 175–178, (2013).
- [72] Birk, Max, and R. L. Mandryk, “Control your game-self: effects of controller type on enjoyment, motivation, and personality in game,” *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 685–694, (2013).
- [73] Shafer, M. Daniel, P. Corey and L. Popova, “Spatial presence and perceived reality as predictors of motion-based video game enjoyment,” *Presence: Teleoperators and Virtual Environments*, 20, no. 6, pp. 591–619, (2011).
- [74] J. Ahmad, S. Kamal and D. Kim, “Shape and motion features approach for activity tracking and recognition from kinect video camera,” *In International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 445–450, (2015).
- [75] Piyathilaka, Lasitha, and S. Kodagoda, “Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features,” *In 8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 567–572, (2013).
- [76] J. Ahmad, Y.H. Kim, Y.J. Kim, S. Kamal, and D. Kim, “Robust human activity recognition from depth video using spatiotemporal multi-fused features,” *Pattern recognition* 61, pp. 295–308, (2017).
- [77] J. Ahmad and Y. Kim, “Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data,” *In International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 119–124, (2014).
- [78] J. Ahmad, S. Kamal and D. Kim, “A depth video sensor-based life-

logging human activity recognition system for elderly care in smart indoor environments,” *Sensors* 14, no. 7, pp. 11735-11759, (2014).

- [79] J. Ahmad, S. Kamal and D. Kim, “Individual detection-tracking-recognition using depth activity images,” In 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 450-455, (2015).
- [80] Wu, Haitao, W. Pan, X. Xiong and S. Xu, “Human activity recognition based on the combined svm&hmm.” In International Conference on Information and Automation (ICIA), pp. 219-224, (2014).
- [81] <https://www.github.com/victordibia/skyfall> (Accessed on 15 Nov. 2018)
- [82] <http://www.freeaddinggames.com/game/beads-puzzle/> (Accessed on 15 Nov. 2018)
- [83] 2d platformer - asset store. <https://www.assetstore.unity3d.com/en/#!/content/11228> (Accessed on 15 Nov. 2018)



Muhammad Raees

Muhammad Raees is currently serving the department of Higher Education, KPK, Pakistan, as an Assistant Professor of Computer Science. He did M.Sc.(Computer Science) from Peshawar University, Pakistan in 2003. He completed M.Phil. with distinction in Virtual Reality Programming from University of Malakand in 2015. He is pursuing Ph.D. (Virtual Reality & Intelligent Systems) from University

of Malakand. His research interests include Virtual Reality Programming, Computer Vision and Machine Learning.



Sehat Ullah

Sehat Ullah has been serving the Department of Computer Science and IT, University of Malakand Pakistan, since 2001. Currently he is Associate Professor of Computer Science and HOD of the Virtual Reality Department. He has Ph.D. degree in the field of VR and 3D Interactions from University of Évry, France. He has supervised more than ten Ph.D. scholars and has about 30 research

publications to his credit.