Vision Based Human Gesture Analysis for Human Machine Interactive Applications

Triveni P. Kulli, Maheswari A.

Abstract- ***In human-Machine Interaction (HMI), hand gestures provide an interactive natural communication between human and machines. Specifically, the visual interpretation of hand gestures will improve the usability, efficiency and quality of HMI. Recently, researchers are more interested in exploring computer vision-based analysis and interpretation of hand gestures, especially after the release of Microsoft Kinect depth sensor. Major applications of human gesture analysis are surveillance, medicine, enhanced visual games, man-machine interface and animation. Using the low cost and high reliable sensor 'Kinect' with its SDK tool kit for windows gives a way to solve with ease some of the difficult problems encountered when working with conventional cameras. The main objective of this project is to control the Human Machine Interactive applications such as graphical user interface, translating of sign language and stroke rehabilitation using Human Gesture Analysis techniques using the Microsoft's Kinect sensor and LabVIEW, a powerful software tool from National Instruments.***

Keywords: Human gestures, Kinect Sensor, Skeletal Analysis, NILabVIEW, Data Acquisition

I. INTRODUCTION

Human Gesture means a movement of a limb or the body as an expression of thought or feeling and it can be called as several names like human pattern, posture, pose and behavior. Human gesture recognition from video streaming has been predominantly studied because of advanced applications such as to enhance monitoring of patients for fall movement detection, motion analysis in sports, and human's various behavioral analysis.Human Gesture is a non-vocal communication which can be used instead of or in combination with a verbal communication, intended to express meaning.The language of gesture is rich in ways for individuals to express a variety of feelings and thoughts, from contempt and hostility to approval and affection. Human gestures are expressive, meaningful body motions involving physical movements of the fingers, hands, arms, head or body with the intent to convey meaningful information or to communicate with the environment. With the rapid development of computer technology, human computer interaction has become ubiquitous activity in our daily life. More attention has been focused on translating these human gestures into computer-understandable language. In the past few years, many gesture tracking and recognition technologies have been proposed.

Revised Version Manuscript Received on June 20, 2015.

Triveni P. Kulli, Student, M.Tech, Electronics and Communication Engineering Sambhram Institute of Technology, M.S Palya, Bangalore-560097, Karnataka, India.

Maheswari A, Asst. Prof., Electronics and Communication Engineering Sambhram Institute of Technology, M.S Palya, Bangalore-560097, Karnataka, India.

Gesture recognition development has been successful with the use of Microsoft's Kinect Sensor, which provides a user to use gestures that are typically interactive and relatively simple to perform various tasks such as controlling games, starting/stopping a movie etc. The Kinect sensor is aninput device which senses the human gestures by capturing the images at faster rate. It was originally developed in November 2010 for use with the Xbox 360 gaming console but has recently been used vastly with Windows PCs for commercial purposes. Kinect camera is selected for human gesture recognition because of its capability to capture depth when compared with other conventional cameras. The Kinect device is composed of multiple sensors. The default RGB video stream uses VGA resolution (640 × 480 pixels) of 8 bit with a colour filter, but the hardware has the capability of resolutions up to 1280x1024 (at lower frame rate) and other colour formats such as UYVY. The Kinect is a fast and accurate optical sensor for extracting 3-Dinformation from 0.4m to 7.0m compare with other optical depth sensors. Since the Kinect camera is using optical sensor, non-optical sensor have to be used to assist "seeing" the crystalline or highly reflective objects. Though the Kinect sensor is vastly used in facial, voice and gesture recognition, the sensor is still under various research of development for security purposes. LabVIEW (LABoratory Virtual Instrumentation Engineering Workbench) is a platform and development environment for a graphical programming language from National Instruments. LabVIEW is a graphical development environment for creating advanced custom applications that interact with real world data, events or signals in fields such as science and engineering technology.What LabVIEW makes it as a unique software tool is that the availability of wide variety of tools in a single environment, ensuring that compatibility is simple as drawing wires between the functions.LabVIEW with the key features like Data Acquisition, Data logging, Data Analysis and Visualization toolkits, makes the researcher to develop the gesture analysis applications at the faster rate with more flexibility and features incorporated.

II. LITERATURE SURVEY

Gesture Tracking is a technology that achieves the human-machine interactions without requiring of physical touch or contact based inputs to the system. The main objective of gesture recognition involves creating a system capable of interpreting specific human gestures using various mathematical algorithms and using them to give meaningful information or to control the devices [1]. For gesture recognition, high recognition rate can be obtained by independently using inertial sensors or vision sensors, especially when real-time recognition is not required. But for real-time gesture tracking, inertial sensors suffer from the zero-drift problem while vision sensors have poor performance for resolving fast motions due to motion blur and occlusions. Hence, neither of them is perfect for gesture tracking alone. Hybrid gesture tracking based on vision and inertial sensor fusion offers not only tracking the fast movements and good stability, but also robust performance [11]. Gesture trackinghas a huge range of real-world applications, such as augmented reality [12], surgical navigation [13], and ego-motion estimation for robot or machine control in industry [4].For all the DOF, (degree of freedom)of the human body, we seem to use a normally a small number of motions that they permit. Even athletes, use their bodies in ways that many people do not use, aspire to repeat motions flawlessly, spending on a lot of hours on practicing the same repeated actions or motions. Most of us use only a subspace of actions or motions out of the space allowed by human’s body Degree of Freedom. For an instant, if the body is modelled by a series of various joints and angles between them, there may be as many as combinations of joint angles that we will never see. Gestures are definitely more interesting subspace of human body movements or motion.The gestures are defined to be motions of the body that are intended to communicate to another person. Hence, the person giving the gesture inputs and the viewer or the receiver of the gesture must have knowledge of gesture to communicate effectively. On simple notes, this means that the gesturer’s motion must be one of a predefined set of gestures. This doesn’t restrict the given gesture that it should have a fixed geometry. For an instant, a “swipe bottom” may be a single gesture in which the hand can be moved from chest to bottom, towards the waist at any height and the angle between them. The system level understanding of human motions and gesture brings out new possibilities to computer human interaction applications. Such interests have inspired research into the stud and the recovery of the complete 3-Dposture of the body or hand using a 3-dimensional physical model. To build such applications, the complete kinematic model of the body is required to have successful interactions.The person providing gesture inputs is called as gesturer. If a gesturer provides the movements to the viewer or receiver of the gesture, the task of gesture understanding is particularly meant to a view based, multiple model approach in which only a small subspace of human motions is represented.The important requirement for gesture recognition is the tracking technology used for obtaining the input data.

Skeletal Tracking

The skeletal tracking has been performed using Kinect combined with the NUI library as described in reference paper [3]. Even though the Microsoft claims that the Kinect can recognize up to six users in its field of view, many researchers found that it can track maximum of only two users that are tracked in detail. The Kinect sensor itself cannot recognize people; it just captures the images at the preset resolution and sends the depth image to the host device, like Xbox 360 gaming consoleor PC/Laptop. The Software installed on the host device will be running the logic to decode the information or the data and recognize elements in the image with human gestures. The software will be preloaded with a wide variety of human body gestures. It uses the alignment of the various body parts, along with the way that they move, to identify and track the various gestures. The application built with the feature of skeleton tracking, can locate twenty, 20human skeletal joints of a user standing and ten upper-body joints (elbow, shoulders, arms, head and wrist) of a user sitting exactly in front of Kinect device. The Kinect sensor is a device build by Microsoft, with the technology of depth sensing, an infrared emitter, a built-in colour RGB camera, and a microphone array. These hardware’s in the device enables the Kinect sensor tosense the location and human body movements along with few voice commands. The latest Kinect sensor V2 is three timesbetter in depth fidelity, than the Kinect 360sensor.It also provides significant improvements in visualizing smaller objects, minute gestures and fast tracking ability.

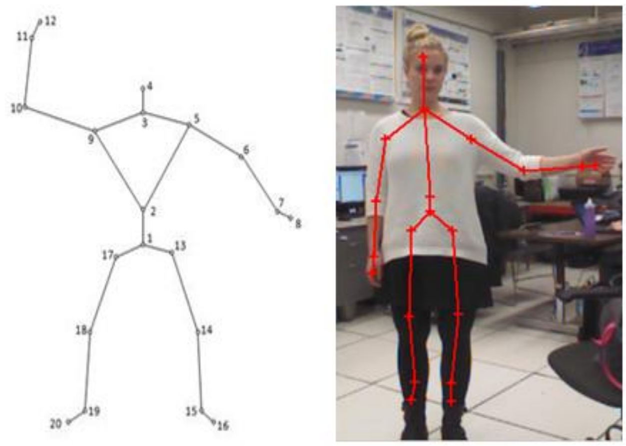

Figure 1: Twenty - Joint point recognition

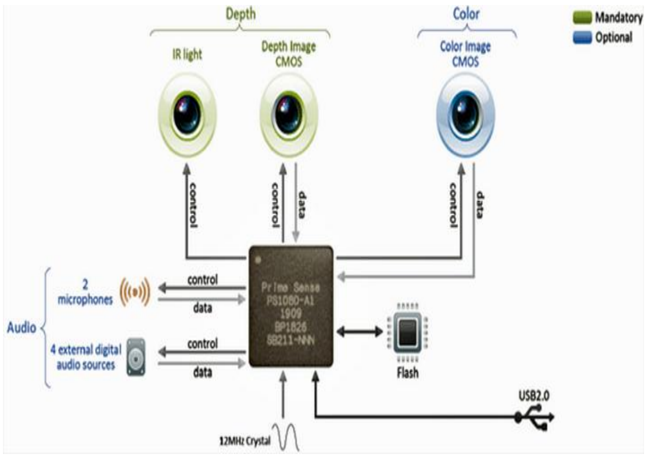

Figure 2: Kinect Sensor Description

The Kinect sensor in incorporated with various advanced sensors for Microsoft's Xbox 360 that provides a controller free gaming and real time experience. It has a color CMOS sensor, five microphones and a standard CMOS image sensor which reads the IR coded light back from the scene.

The Prime Sensor uses the technology called prime sense technology, which enables the three-dimensional capturing capability of the Kinect sensor.These 3 dimensional images are translated in to the depth image. The Prime Sensor is made up of a chip Prime Sense PS1080 SoC, 3D depth sensors and a RGB camera to sense the color image. The Prime Sense PS1080 SoC captures the depth image by directing invisible infrared light on the persons standing in front of the Kinect view. The CMOS image sensor then reads the coded light back from the scene. Then, the Prime Sense PS1080 SoC uses a mathematical algorithm to process the data from the CMOS sensor to create an accurate depth image per frame of the scene.

III. DESIGN AND IMPLEMENTATION OF HUMAN GESTURE ANALYSIS

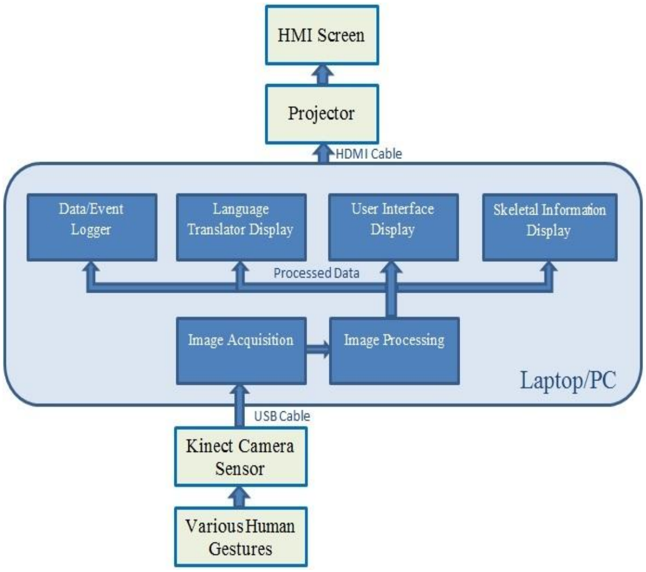

Figure 3: Block Diagram of the project

With the help of Microsoft Kinect sensor, the various human gestures are captured. The captured 3D images are acquired by the processing unit (Laptop/PC) using LabVIEW APIs. These LabVIEW APIs are built using the Microsoft Kinect SDK for Windows. After the image acquisition, the images are processed by applying various algorithms such as Language Translator, Skeletal Joint Tracking algorithms and etc. using NI Vision toolkit. Based on the recognized gesture, the graphical user interface screen will respond to the corresponding actions. The skeletal joint movements are displayed on the screen and the corresponding gesture is translated to the textual language (English). These processed data is then sent to the various displays in which the graphical user interface part is having the interactive controls on the screen, skeletal display will mimic the skeletal movements of the person and language translator display shows the translated data. The language translator display is displayed on the processing unit monitor (i.e. Laptop or PC monitor). The graphical user interface and the skeletal information display will be projected on the HMI screen using a projector. Also during the test is in progress the software on the processing unit will keep on logging the events or the gestures which is given as inputs and will be helpful in post analysis of the test. The graphical user interface screen provides the user to start and stop logging via different gestures. The various human gestures that will be captured are ‘return to home page’, ‘scroll the screen left’, ‘scroll the screen right’, ‘start/stop data logging’, and etc. The above described gestures are captured as a basic functionality and many more such gestures and its responses can be determined. The above figure depicts the block diagram of the project. The image is acquired using Kinect sensor into the LabVIEW environment. The Kinesthesia APIs are used to connect and acquire the images. The acquisition rate is defined such that 16 frames per second are captured and this will give lead to video as the acquisition rate is high. The acquisition rate can be increased beyond 16 frames per second by reducing the loop rate of the while loop. But this requires the high PC/Laptop configuration to support the faster acquisition and also the Kinect sensor specification to be able to capture the images at that rate. The below figure describes the top level software design flow. The Kinesthesia API provides high level of features to initialize, configure and access the Kinect Sensor.

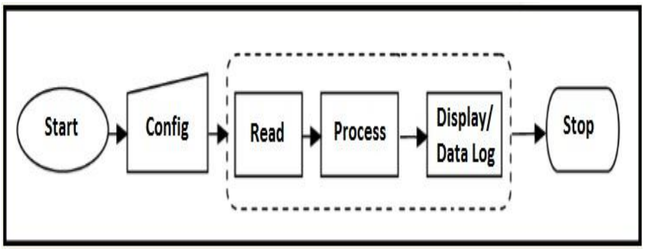

Figure 4: Image Acquisition Process

The above figure shows the flow chart of the image acquisition and processing. The communication with Kinect Sensor is opened using the Initialize API of Kinect sensor. Once the communication is successful then the settings will be done using the configuration API. Here in this, we are configuring it to acquire “Depth and Skeletal image” from the Kinect. The various displays like, graphical interface, textual display, skeletal display are initialized before starting the acquisition process. The Images are acquired at faster rate that is 16 frames per second. This provides an illusion of video acquisition and based on the algorithm incorporated in the Kinect API, the API provides the skeletal information in the form of picture. This is given to the picture indicator of the display and hence the plot of skeletal information is obtained.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

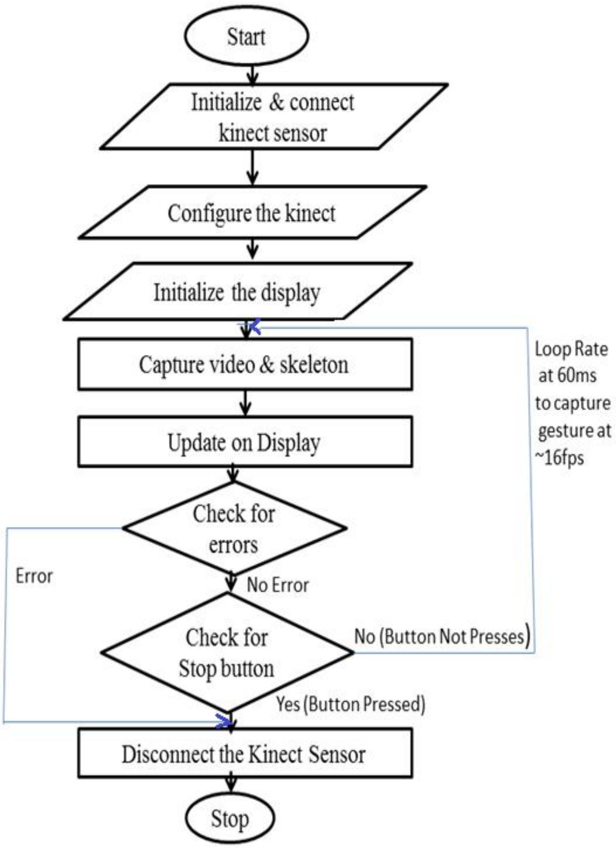

Figure 5: Process Flow Chart

The skeletal information display shows the acquired video and the skeleton of the person. The figure 6 shows the video acquisition and skeletal display. With the help of this display the physiotherapists will be able to predict the movement of the joints and its behavior for patients affected with stroke.

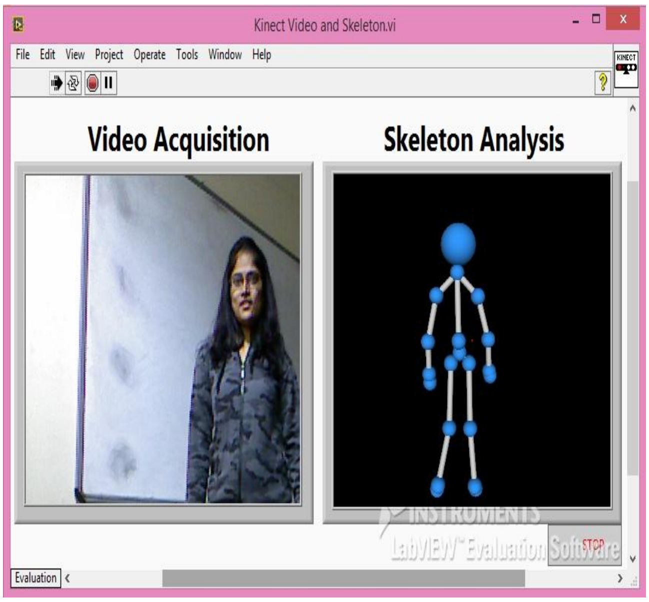

Figure 6: Video and Skeletal Display

The below shown data in the screen are also simulated data of an aircraft engine. The purpose of this display is to provide user the navigation into different screens. This screen also depicts the powerfulness of LabVIEW software in creating such attractive, flexible GUIs. We can customize the controls and indicators as per the requirement and use on the GUI in simple way compared to other GUI based language.

Figure 7: Data View Display

The simulated data and the gesture inputs are logged in the text file. This information's are useful for offline analysis to track the data that are acquired and the user inputs to the applications. This feature finds significant use in the production units where the tracking of user must be monitored. The logged data/ events are with respect to the timestamp. The date of the file created is also based on the system date and the time so the duplication or over writing on same file and data is avoided.The start and stopping of logging is user controllable via a button as well as the gesture. On swiping top gesture, the logging will start and giving the push gesture, the logging will stop.

When the user once again does the swipe top gesture, the new file with current date and time of the system will be generated and the data and events eventually starts logging in to it.

IV. EXPERIMENTAL RESULTS

The skeletal information is successfully tracked and analyses of its movements are observed. The skeletal tracking has been carried out for more than 10 persons and analyzed that the skeletal tracking behavior of the system response is unchanged. The figure 8 depicts the skeletal information tracking. The figure 8 to figure 12 shows the capturing of different gestures and are being converted into Text Language.

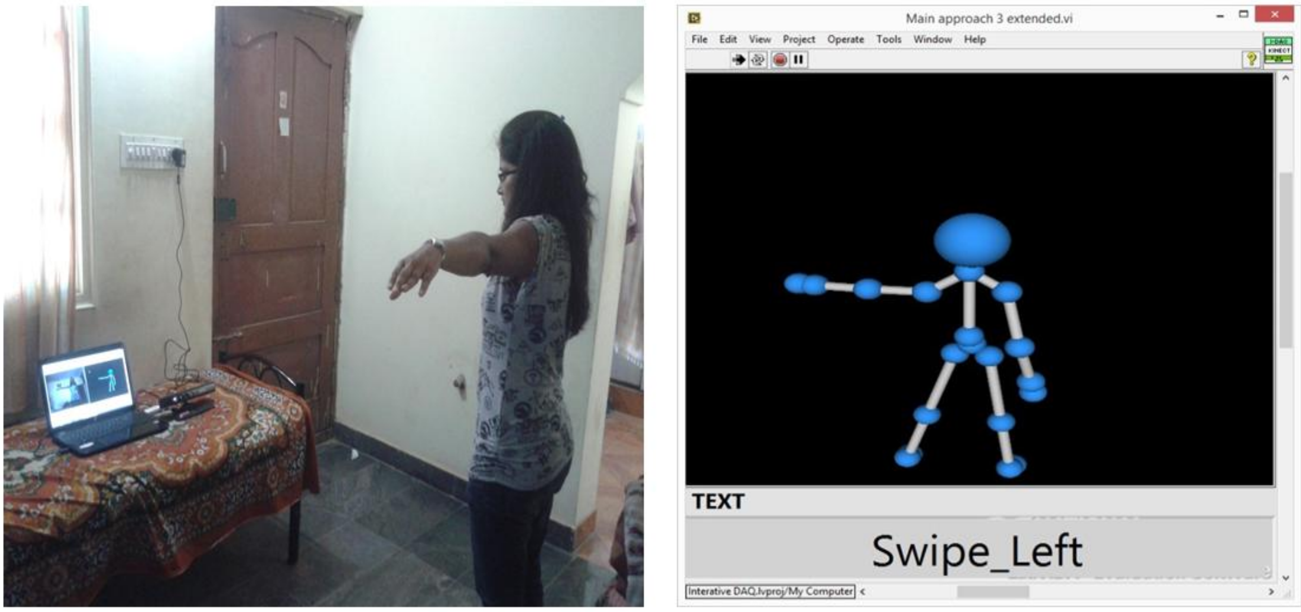

Figure 8: Swipe Left

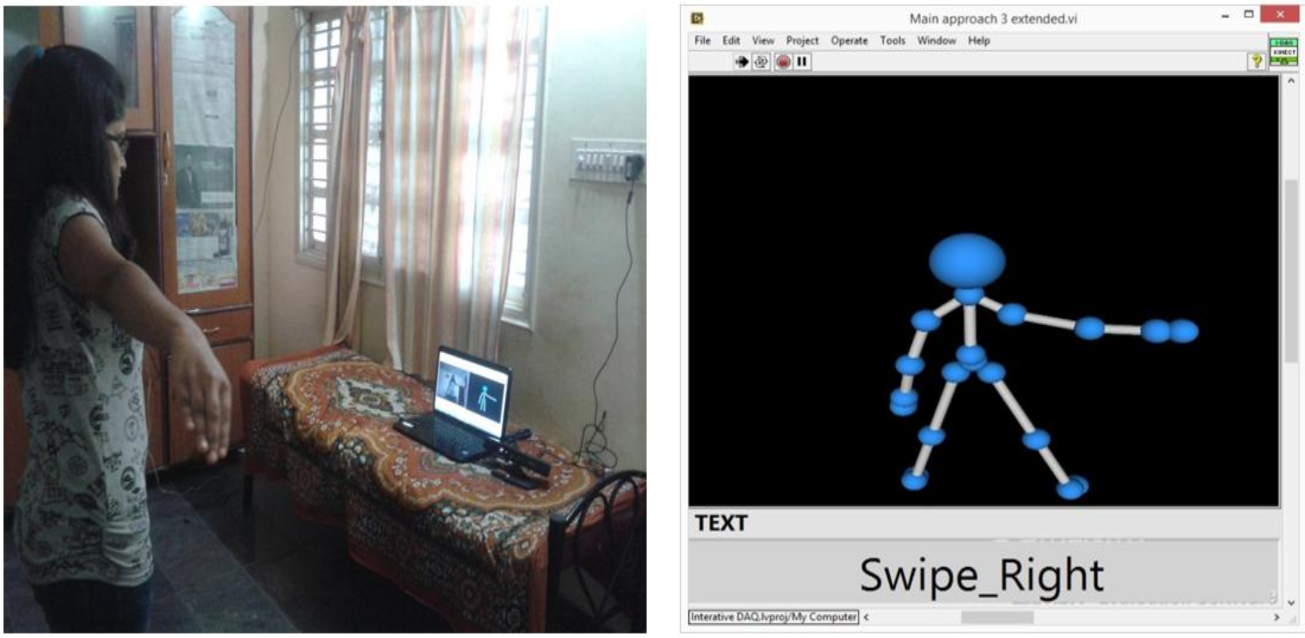

Figure 9: Swipe Right

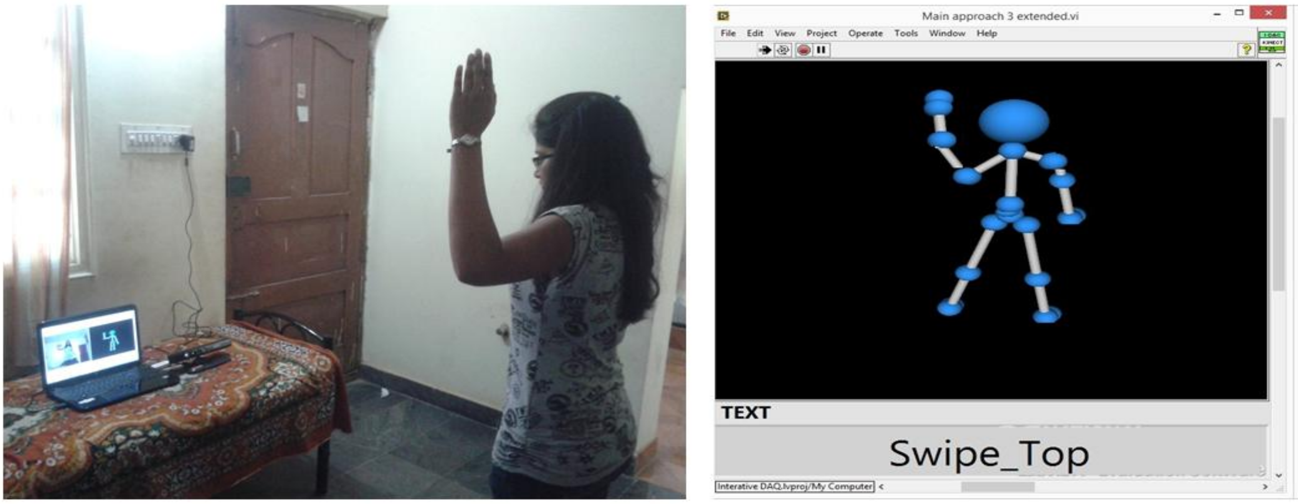

Figure 10: Swipe Top

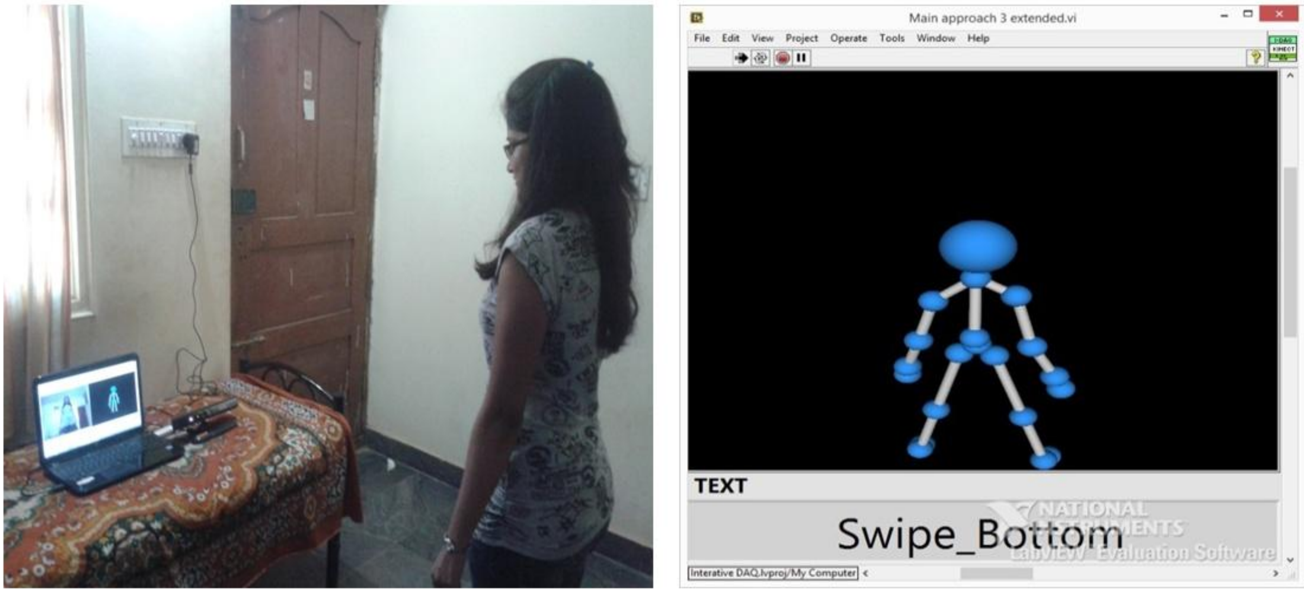

Figure 11: Swipe Bottom

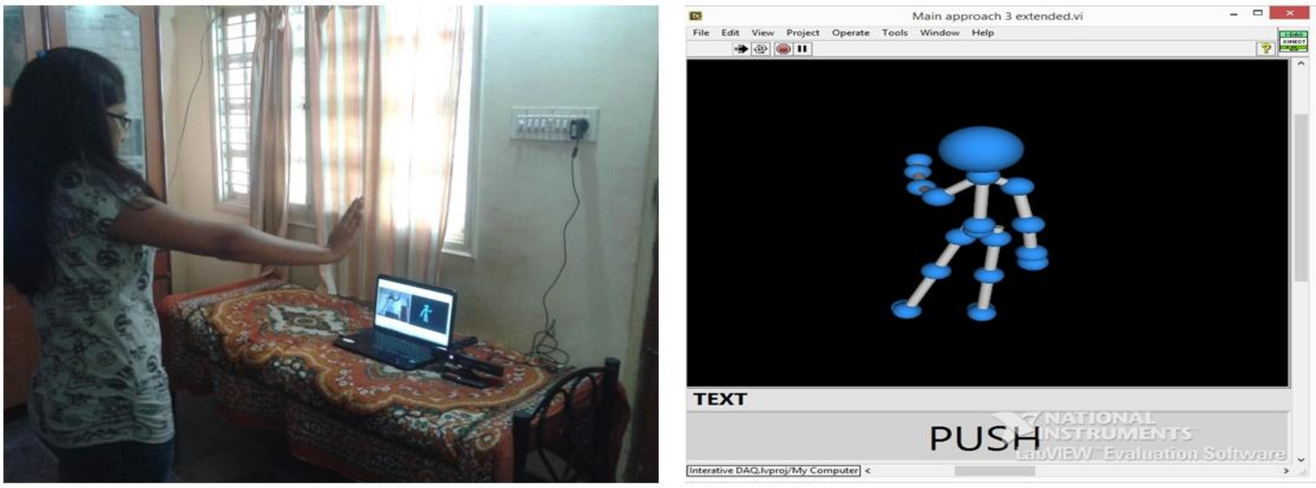

Figure 12: Stop or Push Gesture

V. CONCLUSION

The recognition of various human gestures has been analyzed and based on the recognized patterns; it is translated to corresponding textual language (English). The various actions are performed on graphical user interface viz. swiping left, right, top bottom and etc. The skeletal joint information is tracked and displayed on the HMI screen.

This interface makes the users to be able to control the interactive environments by body orskeleton gestures. The custom graphical interface built using NI LabVIEW software makes the user to navigate to different screens with the pre-defined gestures which is captured by Microsoft Kinect Sensor. The table 1 shows the analysis of control over user interface based on human gestures.

Table 1: Interactive GUI and Gesture Analysis

Gesture	Graphical User Interface
Return to Home	Returns to home screen
Scroll or Swipe Left	Move the screen to left
Scroll or Swipe Right	Move the screen to Right
Scroll or Swipe Top	Move the screen to Top and start data logging
Scroll or Swipe Bottom	Move the screen to Bottom
Stop or Push	Stops data logging

ACKNOWLEDGEMENT

A heart full thanks to our beloved parents, Prof. Ravishankar C.V, HOD of the Department of Electronics and Communication Engineering and also Asst.Prof, A.Maheswari, Department of Electronics and Communication Engineeringfor their continuous guidance, support and motivation.

REFERENCES

1. Toyin Osunkoya1, and Johng-Chern Chern2 Department of Mathematics and Computer Science Chicago State University Chicago, IL 60628, "Gesture-Based Human-Computer-Interaction Using Kinect for Windows Mouse Control and PowerPoint Presentation".
2. Thi-Lan Le, Minh-Quoc Nguyen, Thi-Thanh-Mai,"Human posture recognition using human skeletonprovided by Kinect", Nguyen International Research Institute MICAHUST - CNRS/UMI-2954 - GRENOBLE INPHANOI UNIVERSITY of SCIENCEand TECHNOLOGYVIET NAM, 978-1-4673-2088-7/13©2013 IEEE
3. OrasaPatsadu, ChakaridaNukoolkit and BunthitWatanapa, School of Information TechnologyKing Mongkut's University of Technology ThonburiBangkok, Thailand, "Human Gesture Recognition Using Kinect Camera" 978-1-4673-1921-8/12©2012 IEEE
4. Shengli Zhou, Fei Fei, Guanglie Zhang, John D. Mai, Yunhui Liu, Jay Y. J. Liou, and Wen J. Li, *Fellow, IEEE, "*Human Gesture Tracking and Recognition by the Fusion of MEMS Inertial and Vision Sensors" IEEE SENSORS JOURNAL, VOL. 14, NO. 4, APRIL 2014.
5. Sungil Kang, Juhee Oh, and Hyunki Hong, Dept. of Imaging Science and Arts, GSAIM, Chung-Ang University Heukseok-Dong, Dongjak-Gu, Seoul 156-756, Korea, "Human Gesture Detection based on 3D Blobs andSkeleton Model" 978-1-4799-0604-8/13©2013 IEEE
6. *Megha D Bengalur,* Department of Electronics and Communication Engineering.BVBCET, Hubli, "HUMAN ACTIVITY RECOGNITION USING BODY POSE FEATURES AND SUPPORT VECTOR MACHINE", 978-1-4673-6217-7/13 © 2013 IEEE
7. Loren Arthur Schwarz, ArtashesMkhitaryan, Diana Mateus, Nassir Navab, Computer Aided Medical Procedures (CAMP), Department of Informatics TechnischeUniversitätMünchen (TUM), Boltzmannstr. 3, 85748 Garching, Germany, "Human skeleton tracking from depth data using geodesic distances and optical flow", 0262-8856/$ – see front matter © 2011 Elsevier B.V. All rights reserved doi:10.1016/j.imavis.2011.12.001
8. Leandro Miranda a, Thales Vieira a, Dimas Martíneza, Thomas Lewinerb,Antonio W. Vieira c,d, Mario F. M. Campos c, a Institute of Mathematics, UFAL, Maceió, Brazil,b Department of Mathematics, PUC–Rio, Rio de Janeiro, Brazil,c Department of Computer Science, UFMG, Belo Horizonte, Brazil, d Department of Mathematics, UNIMONTES, Montes Claros, Brazil
9. Pradeep Kumar B P "design and development of HCI using gesture recognition" has Presented in international conference on emerging innovative technology for a sustainable world (ICEITSW-2013) conducted by Oklahoma state university
10. Pradeep Kumar B P" design and development of human computer interface using svm" presented in International conference on Emerging Trends in Engineering (ICETE-12) 15th and 16th May 2012
11. Y. Tao, H. Hu, and H. Zhou, "Integration of vision and inertial sensor for 3D arm motion tracking in home-based rehabilitation," *Int. J. Robot. Res.*, vol. 26, no. 6, pp. 607–624, 2007.
12. Y. Suya, U. Neumann, and R. Azuma, "Hybrid inertial and vision tracking for augmented reality registration," in *Proc. Virtual Reality*, 1999, pp. 260–267.
13. S. Giannarou, Z. Zhiqiang, and Y. Guang-Zhong, "Deformable structure from motion by fusing visual and inertial measurement data," in *Proc.*
14. Marco Soave, Raffaele De Amicis, " Analysis of Human Gestures in the 3D space to control Multimedia Interfaces", Fondazione Graphitech,978-1-4673-2565-3/12/©2012 IEEE
15. Wan yi,Wang jin,Hu Jingwen,Song Tiangang, " A Study in 3D-reconstruction using Kinect Sensor", 978-1-61284-683-5/12/©2012 IEEE
16. Thi-Lan Le, Minh-Quoc Nguyen, Thi-Thanh-Mai,"Human posture recognition using human skeletonprovided by Kinect", Nguyen International Research Institute MICAHUST - CNRS/UMI-2954 - GRENOBLE INPHANOI UNIVERSITY of SCIENCEand TECHNOLOGYVIET NAM, 978-1-4673-2088-7/13©2013 IEEE
17. Sungil Kang, Juhee Oh, and Hyunki Hong, Dept. of Imaging Science and Arts, GSAIM, Chung-Ang University Heukseok-Dong, Dongjak-Gu, Seoul 156-756, Korea, "Human Gesture Detection based on 3D Blobs andSkeleton Model" 978-1-4799-0604-8/13©2013 IEEE
18. SriparnaSaha, ShreyaGhosh , AmitKonar, *1Electronics and Telecommunication Engineering Dept.School of Bioscience and Engineering* Jadavpur University Kolkata, India, "Gesture Recognition FromIndian Classical Dance Using Kinect Sensor**"** 978-0-7695-5042-8/13 © 2013 IEEE DOI 10.1109/CICSYN.2013.11
19. Zhen-PengBian, Lap-PuiChau, Nadia MagnenatThalmann, School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore 639798 * Institute for Media Innovation, Nanyang Technological University, Singapore 639798, "A Depth Video Approach for Fall Detection Based on Human Joins Height and Falling Velocity"
20. OrasaPatsadu, ChakaridaNukoolkit and BunthitWatanapa, School of Information TechnologyKing Mongkut's University of Technology ThonburiBangkok, Thailand, "Human Gesture Recognition Using Kinect Camera" 978-1-4673-1921-8/12©2012 IEEE