Voice and Gesture Integrated Home Automation System

K. A. S. V. Rathnayake[1], W. K. I. L. Wanniarachchi[2], W. H. K. P. Nanayakkara[3]
Department of Physics, University of Sri Jayewardenepura, Nugegoda, Sri Lanka.[1, 2]
CodeGen International (Pvt) Ltd., Sri Lanka[3]
s.vimukthini@sjp.ac.lk[1], iwanni@sjp.ac.lk[2], hansa.nanayakkara@gmail.com[3]

Abstract

In this paper, we propose a Voice and Gesture Operated home automation system with a central controller that allows disabled people to manipulate household appliances with their own voice and gesture commands. The developed system consists of three main parts, such as Kinect Motion Sensor, Central Controller, and Network Coordinator. The whole system is based on a Wireless Sensor Network with a gateway to communicate with each device-controlling unit, which is fixed in each room or device group. Sensor node houses a Nodemcu and an Opto-isolator relay module where each Nodemcu works as a webserver and a client simultaneously, in order to communicate with the central computer that is allocated with a designated IP address. With this approach, system shows over 80% accuracy with gesture and voice with external noise level.

Introduction

Gesture and speech recognition technologies, a couple of years back, were deemed as futuristic concepts. Recently those technologies have rapidly advanced with software and hardware technologies and have integrated with speech and gesture. Such as Google Home, Amazon Echo, smart home gadgets like Philip Hue, Ecobee thermostat, etc. These technologies will offer the facility for people to enhance their lives comfort. The proposed system is specially focusing not only on normal people but also on people who need an external aid for their lives.

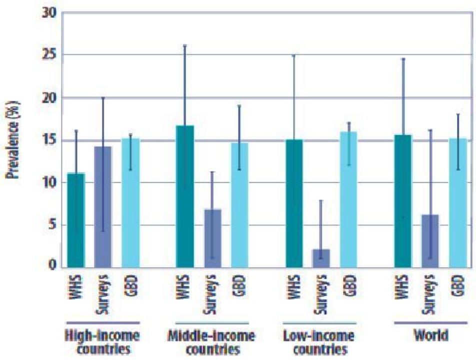


Figure 1. Global disability prevalence estimates from different sources
Source: World Report on Disability 2011

 ISSN 1934-6859

Figure 1 above compares the population-weighted average prevalence of disability for high-income, middle-income, and low-income countries from multiple sources. The solid purple bars show the average prevalence based on available data; the range lines indicate the 10th and 90th percentiles for available country prevalence within each income group. According to the World Health Survey based on more than 100 countries, which was carried out in 2010, there were around 15.6% persons of 15 years or older living with disability. Global Burden of Disease estimates 190 million (3.8%) have "severe disability," the equivalent of disability inferred for conditions such as quadriplegia, severe depression, or blindness including children (about 15% of the world's population) to be living with disability (WHO, World Bank, 2011). According to the World Health Organization's (WHO) reports, global disability prevalence is growing day by day due to many circumstances (WHO, World Bank, 2011). People with disability have ability to live an independent life with satisfaction when there are tools and technology that can assist them in various ways to reduce the gap between them and the society. This proposed system mainly focuses on physically impaired people and its only aspiration is to make them more comfortable with their own room appliances.

Many home automation researches use various communication options to monitor and control devices and appliances such as mobile network, Radio Frequency (RF), Wi-Fi, Internet, LAN technologies and so on. Gill, Yang, Yao, & Lu (2009) proposed a home automation system based on ZigBee which used radio frequencies for communication in between devices and the user. Remote user can access the gateway and connects with internet enabled and Java supportive devices. Local user can access the gateway and controlled devices by using a remote control. But at present almost in every place there is a Wi-Fi router; therefore, the above system will be useless for a place where a Wi-Fi router already exists. Piyare & Tazil, (2011), P. Reddy, K. Reddy, P. A. Reddy, Ramaiah, & Kishor (2016) and Kodali, Jain, Bose, & Boppana (2016) researches introduced home automation systems based on android applications. But their communication methods differ as Bluetooth module based communication and Wi-Fi module based communication. Bluetooth based system contains central Arduino Bluetooth controller and collects data from cell phone, which has a Python script (Piyare et al., 2011). According to their research explanation, Bluetooth communication protocol reduces the cost of the system by reducing cost for wiring, which results single shared Bluetooth module among all the devices, and the disadvantage is that it experiences an access delay. Moreover, its low data rate and shorter data range nature gives less flexibility in to the system. Wi-Fi based systems which were introduced by P. Reddy et al. (2016) and Kodali et al. (2016) provide high data rate and long data range in to the system but user needs another controlling device (mobile phone) to start the communication.

 ISSN 1934-6859

Home automation recently developed systems with voice recognition and gesture recognition, but the recognition accuracy and the feasibility of those systems for the physically impaired person are important factors. Gupta (2015) and Amrutha, Aravind, Ansu, Swathy, Rajasree, & Priyalakshmi (2015) have proposed some systems that use voice recognition to control home appliances and Gupta (2015) used trained HM2007 voice recognition module. Commands are stored in binary form after analog to digital conversion of the input command, and commands are fed through 8 bit data bus to Atmega16 microcontroller in order to execute the task. Amrutha et al. (2015) proposed a MATLAB programming solution for voice recognition part in their system and the system needs to be trained for the speech pattern of each user in order to maintain its high recognition accuracy. Guesgen & Kessell (2012) introduced a system based only on gestures. Gesture controlling system consists of gesture recorder and a gesture recognizer. The recorder uses the WiiRemoteJ3 Java library to capture the gesture. SwingStates4 Java library is used to implement the state machine behavior of the recorder and for the recognition part of the system.

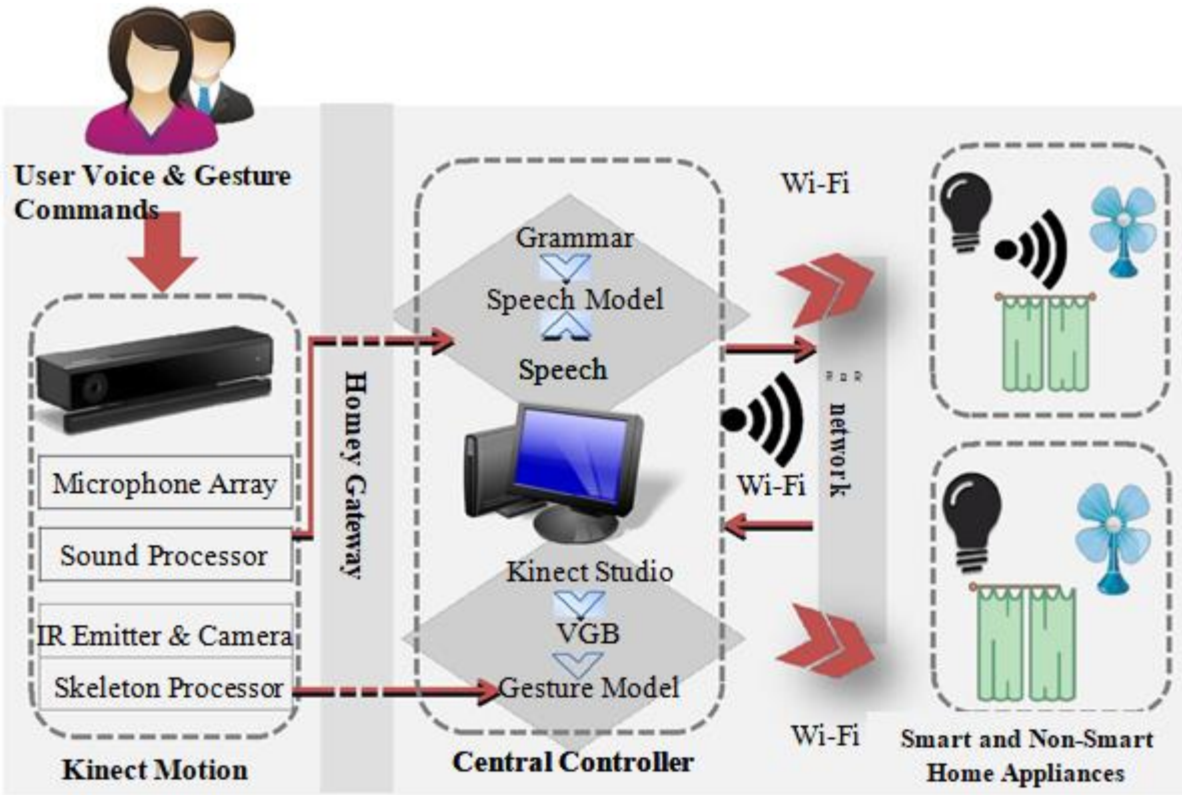
In this research work, we have addressed many difficulties and issues faced by previous home automation systems mentioned above. This proposed system is a low cost, flexible, user friendly, efficient and open source kit for physically impaired people who can use it without any physical touch. It does not require a dedicated Wi-Fi router for communication and the only requirement is having Kinect sensor as the extra hardware device, which is responsible for providing high output accuracy for the system. The assembling and arranging the system does not make any burden on the user in wiring the entire house as it facilitates to work with any plugging location. Moreover, since Kinect utilizes an active Infar Red (IR) sensor for body-movement tracking, the system enables the user to operate it without any visible light. The structure of the System User Interface (SUI) gives the user freedom to communicate with many appliances in many rooms and makes the system more flexible and friendlier.

Implementation

The home automation system implemented focusing on differently abled people basically consists of three main parts such as Kinect Motion Sensor, Central Controller and Network Coordinator. The whole system is based on a Wireless Sensor Network (WSN) with a gateway to communicate with each device-controlling unit which is fixed in each room or device group. The proposed system was implemented with Microsoft Kinect version 2.0 under Visual c# programing language.

System Architecture

Following section describe the conceptual design of an efficient, user friendly and low-cost solution for a home automation system which can be used by physically impaired people



Kinect Motion Sensor

Main sensor used in this proposed system to extract voice and gesture is Microsoft Kinect Version 2.0, a novel device that has many capabilities of giving output with high accuracy. It contains four microphone arrays with 48 kHz, which has the ability for echo cancellation, beam formation, and sound localization to boost the quality of voice recognition of the sensor. In order to acquire the night vision and give the same resolution Kinect utilizes an IR emitter and IR camera. It can capture VGA resolution video as well as depth images at the same resolution. The color streams captured by Kinect sensor are 1920 x 1080 pixels, 24-bit RGB at 30 fps color video, and 512×424 pixels depth and infrared stream. The effective range of this depth camera is 0.5m to 4.5m from the sensor and vertical and horizontal ranges are 60^0 and 70^0 respectively

Central Controller

Audio stream data coming from preliminary processor after been subjected to kinect audio techniques as well as gestural data enter in to the home gateway for further process. The speech model of the system was developed by using Microsoft Speech Grammar and Speech Recognizer. Speech Grammar provides a set of rules and makes the path for Speech Recognizer to get a meaningful output for the vocal input of the user. Recognition engine (Speech Recognizer) bears the built grammar objects and work together to build up the speech model of the system. The speech model provides a confidence value that portrays the certainty of the result and that value lies between zero and one. A confidence threshold value is predefined for the model and compares its value with the confidence value generated by the model at each gesture occurring time and gives the output, if that confidence value is greater than the threshold value. Best suited threshold value was determined by considering the required recognition accuracy.

In this prototype, gesture command database was trained with the help of Kinect Studio and Visual Gesture Builder (VGB). Kinect studio was developed with custom tools in order to record body data and utilize them in VGB, which was used to handle gesture recognition and construct the gesture model. Proposed system uses discreet gesture commands which are distinguished by AdaBoostTriger detection technology that uses Adaptive Boosting (AdaBoost) machine learning algorithm to track the performance of the user. While the recorded commands are being trained, a boolean value that represents the occurrence of a certain gesture during the entire video clip is assigned and that tagging is used to appraise the occurrence of the gesture and determine the confidence value of that gesture event.

Network Coordinator

Following Figure 3 shows the network topology of the system and illustrates the interconnection between gateway and the other system components. The system facilitates to configure the homey system without internal hard wiring and it can be applied for an established house as well as for a newly constructed house.

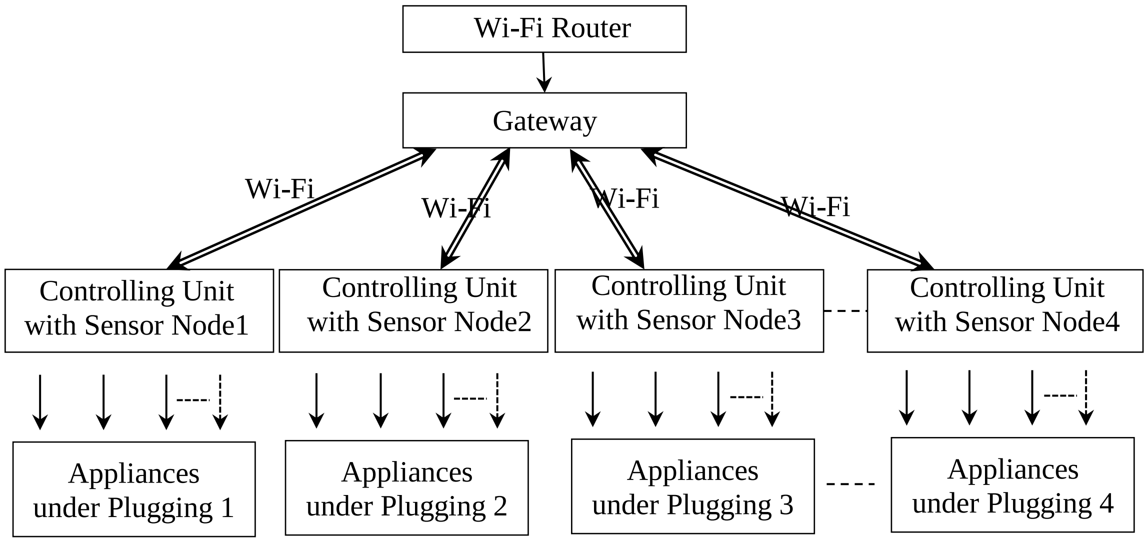


Figure 3. Wireless Network

Wi-Fi based wireless central controller acts as the gateway that coordinate data between its own webserver and the sensor nodes employed with plugged appliances. Sensor node houses a Nodemcu and an Opto-isolator relay module where each Nodemcu works as a webserver and a client simultaneously in order to communicate with the central computer, which is allocated with a designated IP address. Controlling unit then communicate with the non-smart appliances which are connected to the relevant plugging. Smart appliances connected with the system can communicate directly with the Controlling unit without contacting through a plugging module.

NodeMcu

NodeMcu is an open source development board with ESP8266-12E Wi-Fi module which works under eLua based firmware. Embedded Tensilica L106 32-bit microcontroller unit features extra-low power consumption and 16-bit RSIC, reaching a maximum clock speed of 160 MHz (ESP8266 Overview | Espressif Systems).

Graphical User Interface (GUI)

One of the primary objectives of the system is to control myriads of appliances in the home without using any physical touch. In that path to the target GUI of the system bears a huge portion. The system runs under the default mode of voice and it enables the user to switch to gesture mode after devices are connected to the system. First, user or administrator is enrolled with the system after the registration by adding his or her details and the details of appliances in the home; such as, number of rooms, names of rooms and number of appliances, and appliances names in each room.

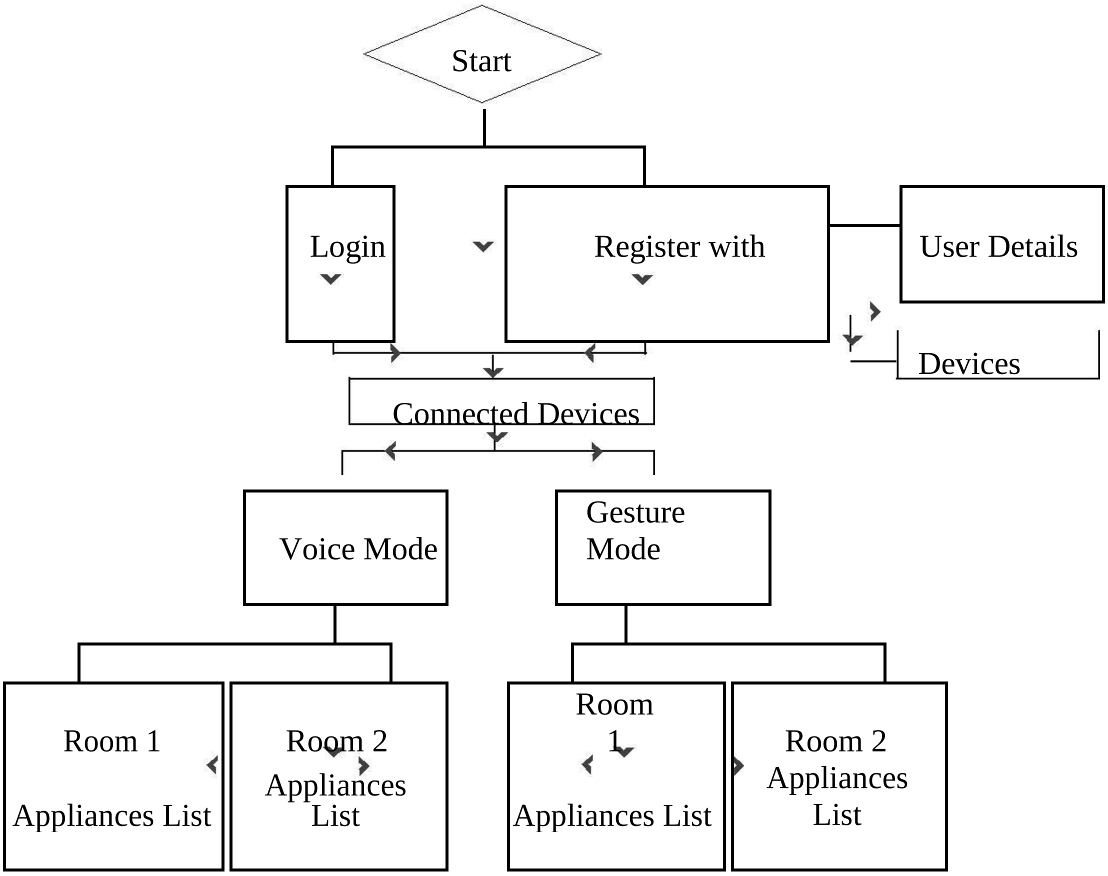
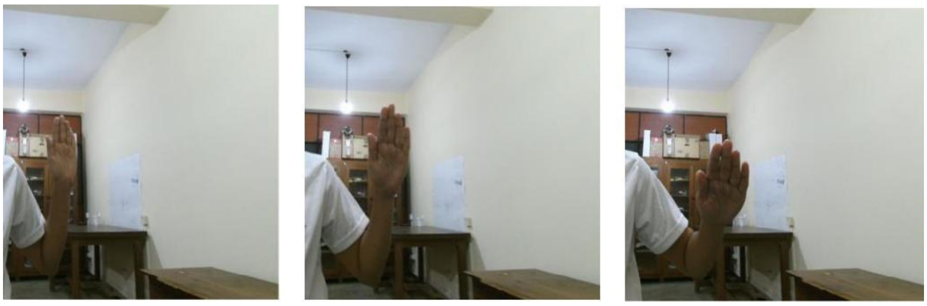


Figure 4. GUI Design

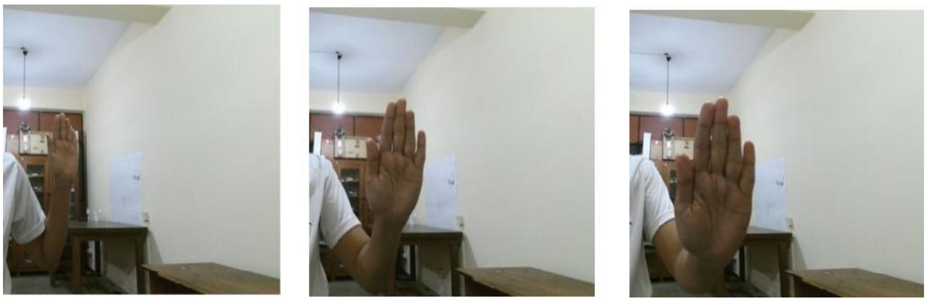
The maximum number of rooms and appliances that can be added by the user are 10 and 10 respectively. All the other following users can be entitled to the system by the registered administrator. With the help of commands displayed in the GUI, the user can follow each step and ultimately can control the specific appliance.

Gesture Commands



Picture 1. Scroll down

Homey UI allows the user to select items from the items list in the room page. Scrolling down the items is done by moving the user's right hand from up to down as shown in the above Picture 1.



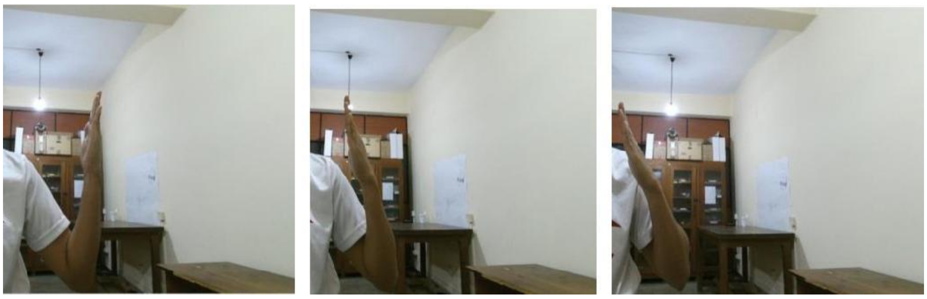
Picture 2. Select

Moving user right hand outwards from the front side of the body enables the item selecting command.

Picture 3. Turn On

Picture 4. Turn Off

Turning on and tuning off the selected item are done by moving right hand and left hand outwards from the side of the body respectively.



Picture 5. Go Back to the Previous Page

Whenever user needs to preview the last page or any previous page user can go back to that desired UI page by moving his right arm towards his body as shown in the figure.

Experimental and Results

Homey system was tested for two major parts as user voice mode and user gesture mode, in order to get a picture of the system accuracy. In our previous research, we introduced a device controlling interface with voice commands for those who are physically challenged. That paper implemented an experiment to test the accuracy of the voice recognition and the system implementation (Rathnayake, Diddeniya, Wanniarachchi, Nanayakkara, & Gunasinghe, 2016).

The test environment was an isolated room and assumed as user bed room with dimension 6m×6m×12m. The experiment was implemented using two parameters. They were performance distance at the specific noise level and the accuracy of Kinect output for the command. The performance distance was taken by measuring the distance between Kinect and the place exactly where the recorder was placed. A recorded command was used for the experiment in order to reduce human errors and its sound level was 100.0 dB ± 0.5 dB. A constant audio sound was used as the experiment noise and it was placed 3 m away from Kinect sensor. The experiment was conducted under the background noise range 47 dB ± 1 dB. The performance distance was increased from 1m to 4 m keeping the recorded audio volume and the noise level constant. The experiment was repeated twice for each distance level to increase the accuracy of the data. The size of a data set was limited to 20 and the program was restarted for each data set as it adapts to the user’s speech pattern. The same experiment was repeated by decreasing the noise level from 100 dB, 95 dB and 90 dB at the noise source (Rathnayake et al., 2016). Noise level I, Noise level II and Noise level III were measured with respect to

Kinect Sensor. Following figure 5 shows the variation of user voice recognition rate with the distance to Kinect from the user.

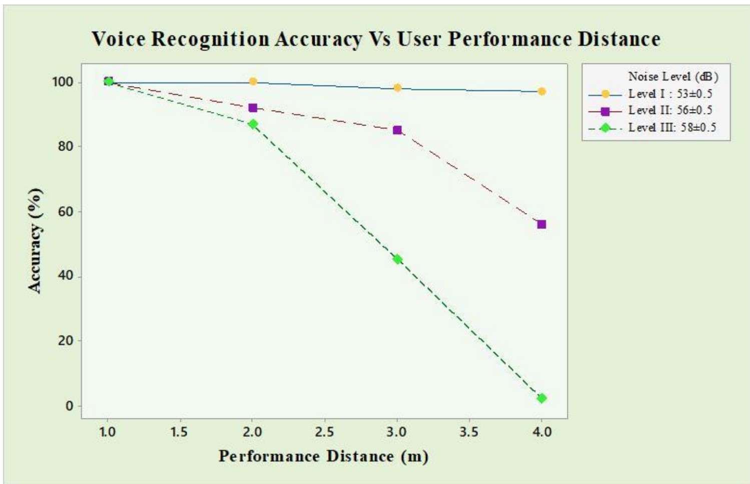


Figure 5. Graph of voice recognition accuracy vs. performance distance

According to the experiment, the recognition accuracy rapidly decreases with the distance when a considerably high noise level is present. But, it shows that when the maximum noise level and the command sound level are in the same distance to the sensor, it still gives 45% accuracy.

To test the system with gesture commands, three people were asked to perform each gesture command 10 times with 3 different speeds and with different distances. The performance distance was varied from 0.6m to 4m by considering Kinect visual range. Among defined five gesture commands, each command was repeated 10 times by starting the performance distance 0.6m in a laboratory room under the luminous intensity 50lux - 53lux. The experiment was repeated for 1.6m, 2.6m, 3.6m as well and recognition accuracy was obtained by varying the speed with three levels as slow(0.5s), medium(1.0s) and high(2.0s). The speed was defined with respect to the gesture performance time and each time-duration was obtained by taking twice of the previous gesture performance time.

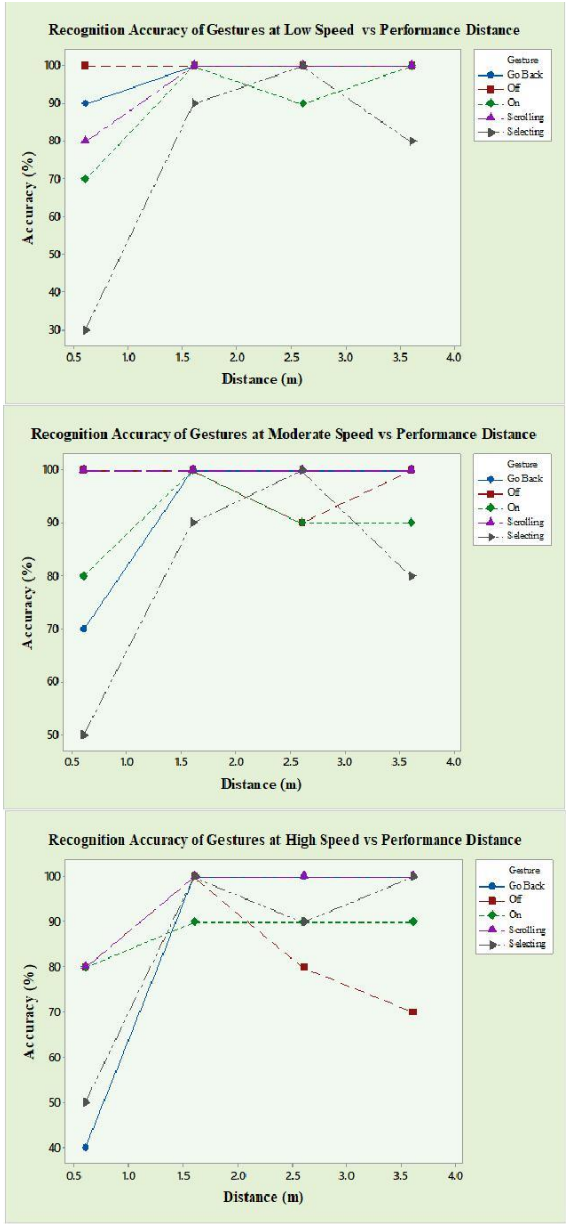


Figure 6. Graphs of gesture recognition accuracy vs. performance distance at each speed level

The line charts depict an upswing of recognition accuracy at every speed until the performance distance is approximately 1.5m. When “Off ” gesture at high speed, it showed more than 60% accuracy at 3.5m, all the other gestures at any speed showed more than or equal 80% accuracy even at 3.5m performance distance. From 1.4m to 3.6 m, the accuracy of every gesture at low and medium speeds was greater than 80%. Altogether, all the gesture commands at every speed showed more than 70% accuracy when the performance distance lay between 1.25m and 3.5m. “Selecting” gesture showed less accuracy at 6m despite the fact that when hand moved towards the sensor at 0.6m, it crossed the Kinect vision range and the output accuracy was violated.

Conclusion

Home automation is becoming a prominent research area in the field of human machine interaction and researchers draw their attention to analyzing, modeling, and recognizing gestures as well as voice recognition when developing a home automation system, since it is the most natural way of communicating and interacting with hardware systems. With that perception, our focus is on self-impaired people who live all over the world and their numbers grow day by day. The performance of the proposed system that predominately addresses those was evaluated by following two designed experiments. According to the experimental results of gesture recognition, the system accomplished well performance by giving gesture recognition accuracy more than 80% at medium and low speeds from performance distance 1.4 to 3.5 m. When the results of voice recognition were analyzed, the system conveyed that, even when the user's sound level and the environmental noise level were equal, still it could have 45% voice recognition accuracy. In conclusion, the system was able to recognize the user speech commands even at a noisy environment and also it was able to recognize gestures not only when the performance distance range was 1.4m - 3.5m and performance speed was slow or moderate, but also when they were fast by 0.5s. Since Kinect utilizes IR for gesture recognition process, the system enabled the user to control his home appliances even in a dark environment.

Future path of the home automation system will cover the issues that have currently immerged in the system. Therefore, in order to fine tune the system, many more people can contribute to train the model further with gesture and voice. It will result in higher accurate recognition than present work. Along with that, we hope to allow the user to customize his own commands with the help of system UI as it furnishes the user system convenience.

Moreover, the target that addresses the security problem of the system will be solved by developing the system with face recognition and entitling the user to log into the system with his face as a security code.

Acknowledgment

Authors wish to acknowledge grant ASP/01/RE/SCI/2016/25 supported for this research work.

References

Amrutha, S., Aravind, S., Ansu, M., Swathy, S., Rajasree, R., & Priyalakshmi, S. (2015). Voice Controlled Smart Home. International Journal of Emerging Technology and Advanced Engineering (IJETAE), 272-276.

Diddeniya, I., Rathnayake, K. A., Gunasinghe, H., & Wanniarachchi, W. K. (2017). Integrating iRobot-Arduino-Kinect Systems: An Educational Voice and Keyboard Operated Wireless Mobile Robot. International Journal of Scientific Engineering and Technology.

ESP8266 Overview | Espressif Systems. (n.d.). Retrieved December 6, 2017, from Espressif Systems - Wi-Fi and Bluetooth chipsets and solutions: http://espressif.com/products/hardware/esp8266ex/overview/

Gill, K., Yang, S. H., Yao, F., & Lu, X. (2009). A zigbee-based home automation system. IEEE Transactions on Consumer Electronics, 422-430.

Guesgen, H. W., & Kessell, D. (2012). Gestural Control of Household Appliances for the Physically Impaired. Twenty-Fifth International Florida Artificial Intelligence Research Society Conference, (pp. 353-358). Marco Island: AAAI Publications.

Gupta, P. (2015). Human Voice Controlled Home Appliances. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 142–143.

Kodali, R. K., Jain, V., Bose, S., & Boppana, L. (2016). IoT based smart security and home automation system. 2016 International Conference on Computing, Communication and Automation (ICCCA) (pp. 1286-1289). Noida: IEEE.

Piyare, R., & Tazil, M. (2011). Bluetooth based home automation system using cell phone. 2011 IEEE 15th International Symposium on Consumer Electronics (ISCE) (pp. 192-195). Singapore: IEEE.

Rathnayake, K. A., Diddeniya, S. I., Wanniarachchi, W. K., Nanayakkara, W. H., & Gunasinghe, H. N. (2016). Voice operated home automation system based on Kinect sensor. 2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS) (pp. 1-5). Galle: IEEE.

Reddy, P. S., Reddy, K. T., Reddy, P. A., Ramaiah, G. N., & Kishor, S. N. (2016). An IoT based home automation using android application. 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 285-290). Paralakhemundi: IEEE.

WHO; World Bank. (2011). WHO | World Report on Disability. Retrieved November 23, 2017, from WHO | World Health Organization: http://www.who.int/disabilities/world_report/2011/report/en/