International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.07, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Automatic control mechanism by Synchronizing Human Machine interface using Computer Vision

Lekha B[1],RabiyaFathima Khader[2], Sowmya M[3],
G Sivananda Reddy[4], Raganna A[5]

[1, 2, 3,4,5]*Student, Department of Electronics & Communication, REVA University(India)*

ABSTRACT

Increasing demand for comfort in personal life has motivated deep research in home automation systems. Several automation techniques in existence utilize sensors and actuators, but the use of video processing system in automation systems would benefit for physically challenged especially the visually challenged for using the automation system effectively. The main theme behind this project implementation lies in automating the process of human activity recognition (HAR) through visual-based recognition systems. While traditional approaches operate on 2-D images and use very computationally intensive algorithms and high dimensional features for activity recognition, recently the introduction of RGB depth cameras has motivated the development of a recognition system with lower dimensional features, the system uses less-complex algorithms and a faster system. The automation system developed here uses a visual recognition system implemented using Matlab. The recognition system uses a Kinect camera as a video capture device and Fuzzy Inference system for making decisions. The Automation system was developed using a hardware setup consisting of Microcontroller unit and Devices to be automated. The complete system consisting of a video capture device, action recognition system and an Automation system was implemented. The proposed Action recognition system is designed to recognize four different actions from a user which is indicated by controlling four different devices. The implemented system shows a real-time performance and suitable for smart home automation systems.

Keywords: *Automation System, Fuzzy Logic, Human activity recognition, Kinect camera, Neural Network.*

I. INTRODUCTION

These days the Human activity recognition and analysis is the most emerging and active concept that has drawn big attention towards improving the human-computer interaction. Human activity recognition concept is based on the proactive computing system that is capable of analyzing various human actions and provides automated features in various real-time applications such as Surveillance System, Health Care System, Traffic Pattern Analysis, Control Entertainment System, etc. there are various approaches for recognition of human activity utilizes the skeletal data information in which the most common approach. The skeleton information is extracted from the motion captured by the kinetic system with the aid of markers positioned on the body. The other techniques extract the spatiotemporal information from the video images, and in this, the recognition process is

based on the large data sets. The utilization of visual automation systems is gaining lots of importance due to their effective features and accuracy. The automation systems have integrated with the several updated and advance feature that are IoT system, Wireless devices, Voice recognition system. Thus, Action recognition is a technique of detecting action from the captured video frames and which involves several processes which are as shown in fig.1 [1] [2].

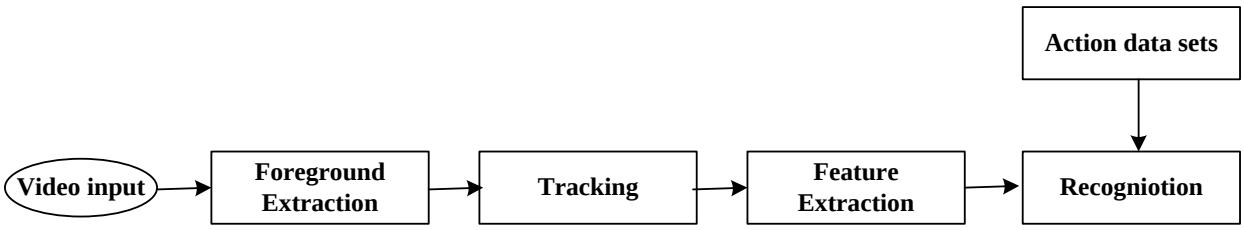

Fig.1. General block diagram of action recognition system

In the activity recognition system the human image object is segmented form the captured video sequences then the object is extracted. An activity classification algorithm is used to recognize the human action behavior. Afterward the tracking process is applied efficiently, and finally, the actions are analyzed by the strength of each video sequences. The existing approaches to Human activity recognition can be classified into three categories such as human model-based approach, a holistic approach and Local Feature extraction approach and a brief description of these approaches are given as follows[3].

a). Human model-based approach: In this, the action behavior is recognition uses a structural feature such as joints position and body parts. In this, the action recognition task is performed by the motion of some the moving light focused on the human body. This approach has a limited scope which is not suitable for surveillance applications and restricted to the indoor environment.

b). Holistic approach: In this approach, spatiotemporal information is directly learned from the frame in video sequences. This method used the information about the localization of body structure in the video and learned an action model that records the characteristic, global body movements without using any information of body parts. The Holistic representation approach is of two types in which the first one uses the shapes mask that originate from the background subtraction and the second one uses shape as well as optical flow information.

c). Local Feature extraction approach: - This approach is based on the local descriptors in which local spatiotemporal feature keeps the information of motion and shapes for the local region in the video. The local method of human action provides many features such as resistance to background differences and avoidance of background subtraction target tracking that required in a holistic approach.

The action recognition has been extensively researched in the past few decades but remains a big challenging factor that comes from difficulties such as great intra-class variance, use of complex machine learning approaches, etc. So to perform optimal recognition task, there is a requirement of the system that computes real-time observation to analyze the human action behavior and can also actuate automation system effectively.

In this project work, an automation system is designed by using a computer vision based HMI recognition system in respect to making control over the home automation device application through action and by using a hand gesture. This paper contains multi fold views that are organized as section 1: describes the general introduction of the human action recognition (HAR) approach, section-2. Focuses on the existing work that related to the human action recognition, section-3 presents the background study followed by proposed system methodology in section-4, section-5illustrates the outcomes of the system designed and finally section-6, reveals the conclusion of the presented work.

II.THEORY

The study of Ofli et al. [4] has focused on the sequence of the most informative joints performance. The sequence of the Most Informative Joints has a very precise practical analysis; in the set of the order in each sequential window informative joints are complete the action in given period. The main motive of this presenting model to show the performance of simple and computationally efficient are much better than the performance of high rated LDS metric models. In practical application, SMIJ is a more prefer human activity recognition task for 3D acquisition.

The work of Jiang et al.[5], have concentrated on a new algorithm K-SVD for the coding of sparse. Its use to learn sparse coding existing discriminative dictionary. In the process of learning of dictionary, the metric column of each dictionary is more efficient in the coding of discriminability sparse. The given optimal solution is capably obtained by the using of presenting this new algorithm. There are presenting some updated coding technique like sparse-coding for face, science, action and given objects are recognized in the category.

The study of Lin et al. [6] ,have concentrated on a new algorithm named HMB for the activity of group recognition; the proposed algorithm have signified the standard of group activities. KPB a new heat-map-based algorithm is proposed for the signify the level of group activities. The last section clarifies the effectiveness of this presenting algorithm.

The work ofPrest et al. [7] has focused on a new source to make easy the learning human-object interaction in videos communication, in which capturing the motion of explicitly is more important by this presenting new source. The main motive is to mix the human-object-interaction features with 3D-HOG to increase the performance of presenting features.

The study of Ji et al. [8] have concentrated on a new 3D CNN approach for recognition of action, its focus on both temporal and spatial dimensions. The updated 3D CNN model was worked along with algorithm and different architecture model. This scheme combination is beneficial in the performance of 3D CNN models on recognized task. The presenting model is implemented in C++ as be a part of human action system. The CNN models belong to the stream of visual recognition.

The work of Cho et al. [9] have described on large motion capture dataset it's based on two kinds of approach one is video and second is motion capture data. The given capture database is a largely controlled dataset it provides the short-cut clip of dataset. There is 130 gesture group in one original dataset, and some motion belongs to a single original class. After completing the action of one dataset, the other is reedy behind one. By using this dataset, they receive the accuracy.

The study of Yun et al.[10] have concentrate on complex human activity dataset its recognized an important field work application like content-based video retrieval, human-computer interface. It's based on MIL classifier in the sequence extend outperforms SVMs. The given algorithm is capable of classifying the short-video by the performance of a periodic action of a single person in this presentation interaction performed two people y the using of human recognize sensor.

The work of Xia et al. [11] have present a scheme to recognize human action according to time and 3D poses. The input is found from the 3D skeletal joint location from the depth of the maps. They developed a new HOJ3D Scheme that presents the posture of human histograms of the location of joint 3D in a coordinate system. The main components of this given algorithm are real-time, and 3D skeletal are included computation of HOJ3D. In the reorganization of human action, the advantage of 3D data also effective of the performance of a task by using of depth information.

The study of Bengio et al. [12] has concentrated on a family of algorithm particularly, like the Restricted Boltzmann machines, components elements. The proposed algorithms are connected with estimators of the gradient in Restricted Boltzmann and Diver-likelihood machines. Its focused on the problem of optimizing the deep design for starting levels of distributed demonstration. By using Restricted Boltzmann technique, they achieved Deep Belief Networks.

The work of Ofli et al. [13] have focused on a multimodal human action database in which largest database more than 80 minutes of data these given data are captured from 12 subjects, in 5 trials and 11 actions by the using of different types of modalities as like multi-view video, acceleration, and audio, mocap. All presenting data are captured by different types of sensor and it's synchronized for temporal alignment.

III. BACKGROUND

This section highlights about background work of related proposed study deals with Kinect sensor device which defines the working role of detecting and capturing the human actions and illustrates about fuzzy inference system. The detail description about those is given in the below sub section.

3.1. Microsoft Kinect Sensor

The Kinect sensor is an advanced version of RGB or depth sensing software and hardware-based technology. In fig: 2 shows the Kinect sensor images where it provides for detecting human faces, synchronize picture signal, capture human 3-Dimensional gesture and others. [14]

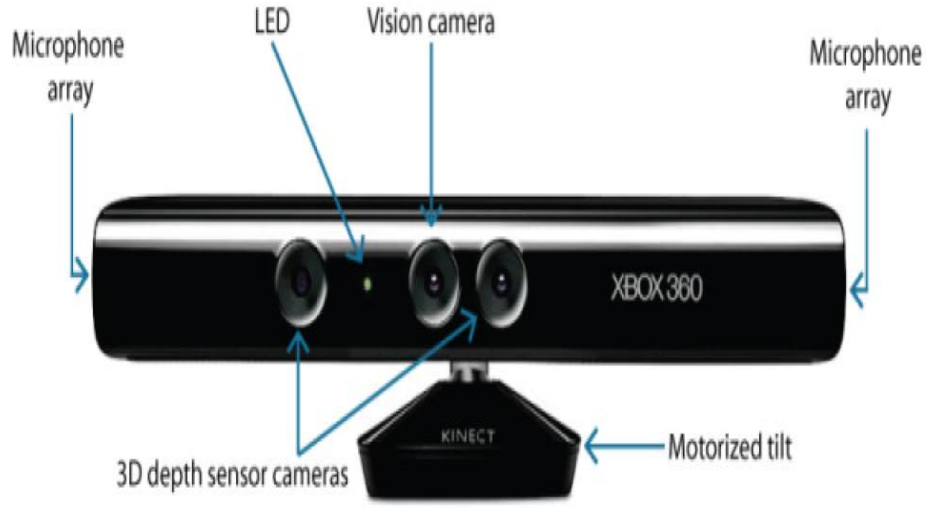

Fig.2. Components of Kinect sensor

Kinect sensor components: the components of Kinect sensor hardware is:

- **RGB camera:** the RGB camera provides three basic colors components of the video. The RGB camera works at 30Hz, and also it can offer pictures at 640×480 pixels. It can also provide HD pictures, which is running at 10 frames/second. The resolution of HD pictures is 1280×1024 pixels.
- **3-Dimensional Depth sensor:** the 3-D depth sensor contain Infrared (IR) laser projector and IR camera. The camera and IR projectors generate a depth map, and it provides the distance calculation between the Kinect camera and object. The sensor range is limit, 0.8m to 3.5m distance is required. The frame rate of output video is 30 frames/second.
- **The motorized tilt:** the motorized tilt is an axis for sensor, and it can be bent up to 27°. It can be in up position or down position.

3.2 Kinect Software Tools

The Kinect sensor needs some essential capabilities in PC for implementation process are: The processor of the PC should be 32-bit or 64-bit, it should be dual-core, 2.66 GHz or more than a faster processor. The PC should have USB 2.0 bus support, to connect the Kinect from PC, and the RAM is minimum 2 GB. The Microsoft Visual Studio 2010/2012 express or different visual studio version is required. Install Kinect Software Development Kit (SDK) to develop the Kinect-enabled application.The SDK in combination with the NUI library that delivers the API and tools.

3.3 Fuzzy Inference System (FIS)

It is an essential unit of fuzzy logic and decision making has a primary work. In developing some necessary decision rules are using, it uses the "IF" or "then" rules and also with connectors "AND" or "OR." The outcomes

of FIS is always in a fuzzy set. A de-fuzzification system unit is used for changing to fuzzy into crisp variables. The FIS is providing a very effective role in data organization, automatic control, and decision analysis.

In below fig: 3 has displays the block diagram of FIS. The functional blocks of FIS have contained some essential units are Rule base, Database, Decision-Making unit, Fuzzification and De-fuzzification interference unit. The rule base consists IF-THEN fuzzy rules. In the database, it described that the fuzzy membership function is used in fuzzy rules. The decision-making unit carries out operations on rules. Next, in the Fuzzification interface unit changed the crisp input into fuzzy quantities and De-fuzzification interface can change fuzzy quantities into the crisp.

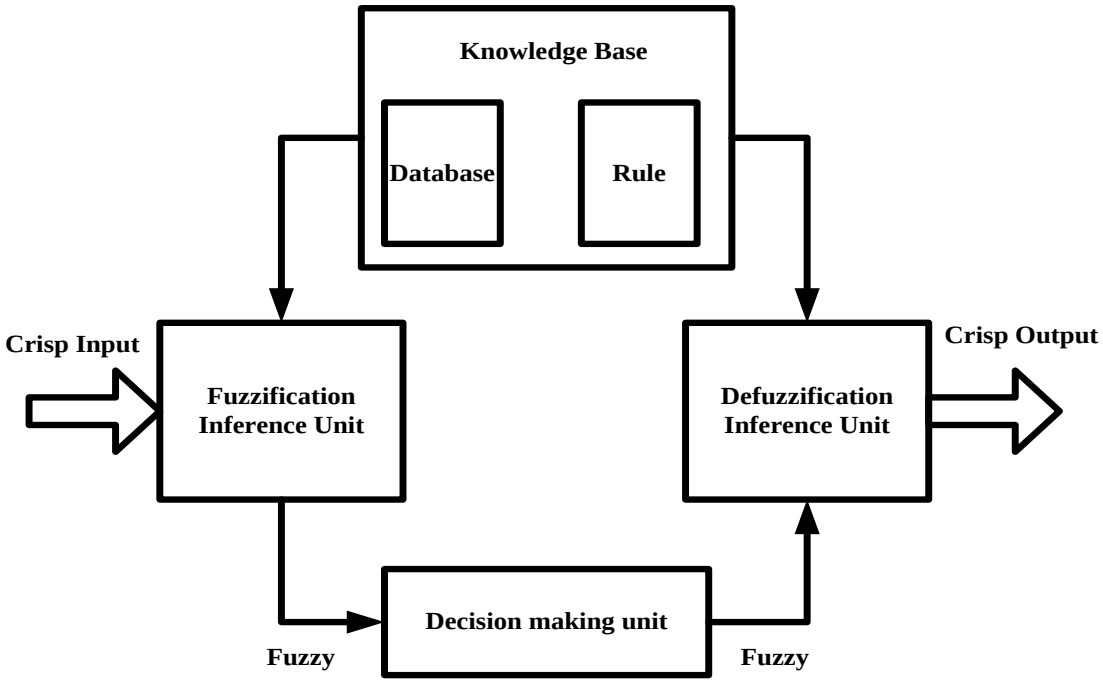

Fig.3.Block diagram of FIS

IV. PROPOSED METHODOLOGY

In this research methodology, have proposed a novel of Hand Video based Human to Machine Interface Recognition (HMIR) system using Fuzzy convolution neural network (CNN). The movement capture tracking information of body joints are utilized to evaluate the temporal changes of displacement among the body joints at the execution of human actions. Thus fuzzy membership operations are considered for feature-extraction which recognizes the human actions. While CNN able to recognize the local action patterns in input-data is trained as well as recognize the actions from the patterns.

Some significant attributes influencing the effectiveness of recognition of human action (RHA) mechanisms are; computation complexity and features considered. Moreover, RHA becomes quite challenging owing to inconsistency in the process of execution of the action and also issue in the

movement alignment during actions recordings. Therefore, to overcome these challenges, the proposed study introduced a mechanism of Human to Machine Interface Recognition system which considers the skeleton motion information of only 3-body joints for feature-extraction and a fuzzy CNN utilized of the classification process. The detail description about proposed computation process is followed in the below sub section.

4.1 Proposed System Block-Diagram

The high-level architecture of proposed HMIR system is presented in fig: 4 (a) along with flow process of the project implementation in fig: 4 (b). According to this, the first system will capture the video, regarding frames with the help of Kinect camera and forwards the color and dept image information to the HMIR system. This recognition system is executed in Matlab code which processes on the host machine.

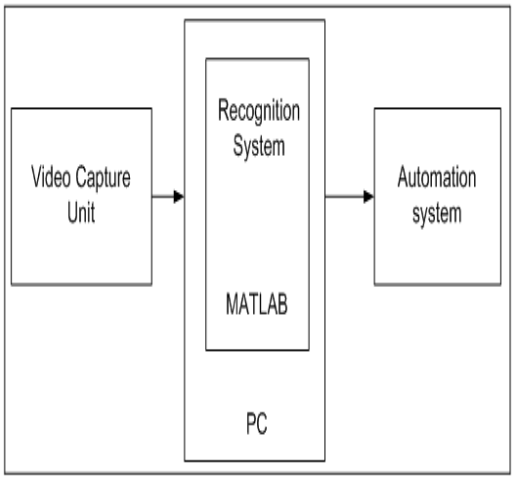

Fig.4. (a) Block Diagram of Proposed HMIR System

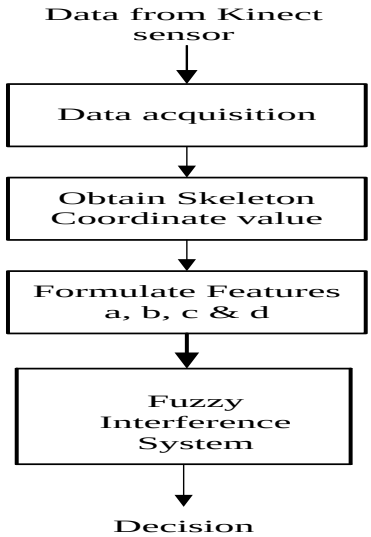

Fig:4 (b) Flow Process of HMIR System

The first system will collect the video motions(i.e., in frames), color, and dept-image information generated from sensor camera (i.e., Kinect camera). Then it constructs the skeleton image of the person. From the coordinates of the constructed image, the system will track the human actions and finds the distance between the coordinates of the point of interest and reference point. Then, the tracked information can be exploited for decision making with the help of decision logics. Additionally, the proposed methodology also adopted a Fuzzy convolution approach (i.e., Fuzzy CNN) for suitable decision making. The logical decision generated from video-based HMIR system is then forwarded to the micro-controller which control the corresponding devices based on decision rule.

4.2 Formulation of motion captures distance measurements.

The information of motion capture for human activity contains 3-D tracking information of skeleton joints during the implementation of action recognition. This information can be utilized for the evaluation of wave angle, distance, and velocity for RHA. In this study, the system only considers the three skeleton joints are tracking information and computes four distance measurements. The skeleton structure of motion capture along with distance measures are given in the blow fig: 5.

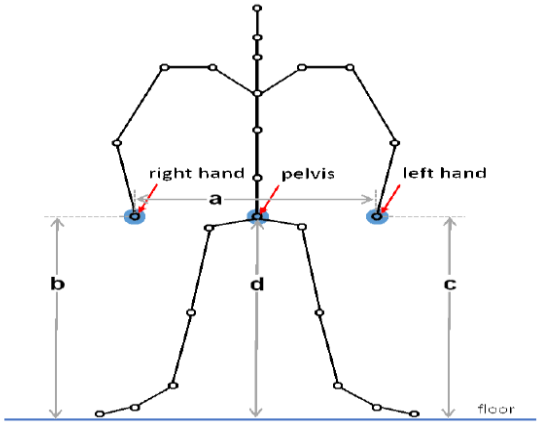

Fig.5. Skeleton structure of motion capture with distance measures

The proposed experimental study considers three different skeleton joints tracking information, i.e., Left Hand (LH), Right Hand (RH) and Pelvis and evaluates the four distance measurements such as 1) distance among the LH and RH, 2) altitude of the RH on top of the ground 3) altitude of the LH on top of the ground and 4) altitude of Pelvis on top of the ground.Here, the system will evaluate the displacement of two specific skeleton joints and utilized for action feature recognition of video-based HMIR system. All four distance measures are generated by considering the distance from the reference point of skeleton joint and point of interest. Here, the RH is the point of interest and shoulder point is considered as a reference point. Therefore, displacement features of RH from the shoulder point are

generated from considering displacement of (a, b, and c) coordinates of RH with (a, b, c) coordinates of shoulder point.

V.RESULT AND ANALYSIS

This section illustrates the outcome results obtained from the successful implementation of the proposed system. The system turns on/off a selected device upon identifying suitable video-based HMIR system.

5.1 Inactive State

In starting phase the system is in an inactive state, and in this phase, it does not identify any movement action because it doesn't recognize the skeleton structure of Human. The below fig: 6 show no activity and also there is no change in automation system state.

Fig.6. Inactive states of the user

Below Fig: 7 displays that the alert system is in its starting state in the inactive position, the system is in deactivate state, and it cannot consider any movement made by the user. That is why the automation system is in an inactive state

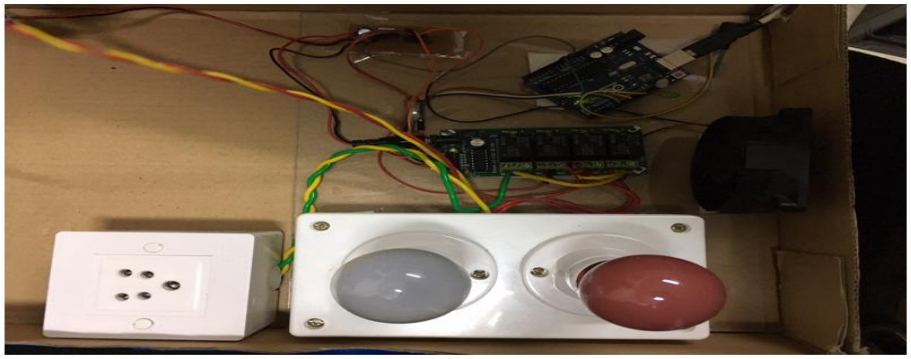

Fig.7. deactivate position of the automation system

5.2 Active state

The automation system is designed such that it turned out to be in an active position after identifying a "Hand Wave" signal from the user.

Case-1Right Hand on Top: In fig 8 shown that when the system is identifying the user action that the "Right Hand on Top" the device two on and fan start rotating.

Fig.8.The snapshot of fan turned on state

Case-2 Right Hand Signal: in fig: 9 displays that the user is moving his right hand to the direction indicated in depth and color pictures.

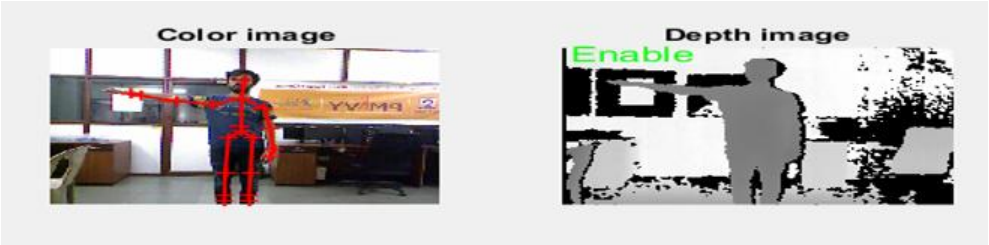

Fig.9. Kinect sensor results for right-hand direction movement

The system identifies the “Right Hand Signal” and turns on the indicated bulb 1 is glowing as displays in fig: 10 below.

Fig.10. Snapshot of bulb 1 turned on

Case-3 Left Hand Signal: in fig: 11 presents the user creating a movement of start moving his left hand to the direction of designated in depth and color pictures.

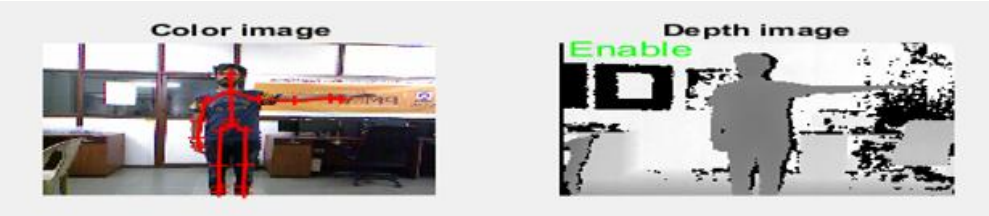

Fig.11.Kinect sensor outcomes for left-hand direction movement

The system detects this action and indicates it by turning on bulb 2. The bulb 2 is glowing as displays in fig: 12.

Fig.12. Snapshot of bulb-2 turned on

The numerical computing platform command windowdisplays the device status as “ON” or “OFF” shown in the fig. 13.

Command Window

Bulb 1 On
Bulb 2 On
Fan on
TV on
Bulb 1 Off
Bulb 2 Off
Fan off
TV off

Fig.13.Command window displays the devices status

VI. CONCLUSION

A Human Activity Recognition system which forms a part of the video surveillance system was designed and implemented. The recognition system was designed based on the Human skeletal features which were obtained using a Kinect sensor. The recognition system was developed as a software system which takes inputs from the Kinect sensor and delivers the detection result to an Automation system. The Automation system was designed as an automation system with four devices controlled for four different actions from the user. The system was successfully designed and implemented on Matlab along with the necessary hardware and software resources. The results showed that the devices were controlled from actions made by the user.The system can be updated to recognize some actions in the future. The system can also be updated to work with 3-D images.

REFERENCES

[1]. Chen, Chen, RoozbehJafari, and Nasser Kehtarnavaz. "A real-time human action recognition system using depth and inertial sensor fusion." *IEEE Sensors Journal* 16.3 (2016): 773-781.

[2]. Xu, Pei. "A Real-time Hand Gesture Recognition and Human-Computer Interaction System." *arXiv preprint arXiv:1704.07296* (2017).

[3]. Ahmed Taha, Hala H. Zayed, M. E. Khalifa and El-Sayed M. ElHorbaty, "Exploring Behavior Analysis in Video Surveillance Applications," In The International Journal of Computer Applications (IJCA), Foundation of Computer Science, New York, USA, Volume 93, Number 14, pp. 22-32. May 2014.

[4] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Sequence of the most informative joints (smij): A new representation for human skeletal action recognition." in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, June 2012, pp. 8–13.

[5] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, November 2013, pp. 2651–2664.

[6] W. Lin, H. Chu, J. Wu, B. Sheng, and Z. Chen, "A heat-map-based algorithm for recognizing group activities in videos." *IEEE Transactions*
on Circuits and Systems for Video Technology, vol. 23, no. 11, November 2013, pp. 1980–1992.

[7] A. Prest, V. Ferrari, and C. Schmid, "Explicit modeling of human-object interactions in realistic videos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, April 2013, pp. 835–848.

[8] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networksfor human action recognition."*IEEE Transactions on Pattern Analysisand Machine Intelligence*, vol. 35, no. 1, January 2013, pp. 221–231.

[9] K. Cho and X. Chen, "Classifying and visualizing motion capturesequences using deep neural networks," *Computer Research Repository(CoRR)*, vol. abs/1306.3874, 2013.

[10] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras,"Two-person interaction detection using body-pose features andmultiple instance learning." in *IEEE Computer Society Conference onComputer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE,June 2012, pp. 28–35.

[11] L. Xia, C.-C. Chen, and J. K. Aggarwal, "View invariant humanaction recognition using histograms of 3d joints," in *Workshop onHuman Activity Understanding from 3D Data in conjunction with CVPR(HAU3D)*, Rhode Island, USA, 2012, pp. 20–27.

[12] Y. Bengio, "Learning deep architectures for ai," *Foundation and Trendsin Machine Learning*, vol. 2, no. 1, January 2009, pp. 1–127.

[13] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeleymhad: A comprehensive multimodal human action database," in *2013IEEE Workshop on Applications of Computer Vision (WACV)*, January2013, pp. 53–60.

[14] Han, Jungong, et al. "Enhanced computer vision with Microsoft Kinect sensor: A review." IEEE transactions on cybernetics 43.5 (2013): 1318-1334.