Current Science International Volume : 06 | Issue : 03 | July- Sept. 2017
ISSN: 2077-4435 Pages: 560-569

Robotic Person Following in a Crowd Using Kinect Sensor

M. S. Zaghloul
Electronics and Communication Department, Arab Academy for Science, Technology, and Maritime Transport, P.O.1029, Alex, Egypt. E-mail: dr_mszaghloul@yahoo.com

Hassan Khaled
Electronics and Communication Department, Arab Academy for Science, Technology, and Maritime Transport, P.O.1029, Alex, Egypt.E-mail: elec.eng.hassan.khaled@gmail.com

Osama M.Badawy
Computer Science department, Arab Academy for Science, Technology, and Maritime Transport, P.O.1029, Alex, Egypt. E-mail: Obadawy@aast.edu

Received: 26 June 2017 / Accepted: 27 July 2017 / Publication Date: 06 August 2017

ABSTRACT

The purpose of this paper is to develop a robotic system that is able to follow a particular person in a real-world environment without losing track due to objects or other people obscuring it. The study focuses on developing image-recognition and motion control techniques for robotic person-following. It uses Microsoft‘s Kinect device, which combines infrared imaging with video imaging for robust recognition, and thus offers the possibility of identifying and following a particular person in a crowd. The proposed system first constructs skeleton images of every person in the frame using the joint coordinate locations provided by an onboard depth camera. To identify its target, it performs pose-recognition based on the skeleton image of the person and learns about the physical characteristics such as height and body-part dimensions. In addition, hue and saturation histograms for the person‘s outfit are used to perform color-based recognition. A significant feature of the study is the introduction of a new histogram matching technique that adapts to changes in environmental light conditions. This, together with skeleton model matching improves the performance of recognition. The second important aspect of the paper is the development of software for the robot that allows the cameras to stay focused on the person of interest. The effort also encompasses research techniques that allow the robot to maintain a relatively constant distance from the person and for graceful and gradual motion transitions for the robot during person-following. The shortcomings of the infrared-based depth sensor and the robotic vehicle limit the system‘s applicability to indoor environments. Nonetheless, the improvements in reliability through the use of the image-processing algorithms developed over the course of this research mean that, if these technical limitations are overcome, the scope of the system‘s applicability will widen significantly.

Key words: tracking, recognition, Arduino , robotic, wireless system

1-Introduction

A key competence of a service robot is the ability to navigate in its human inhabited environment. This requires a map of the environment, Person to track and follow the automated system is an intelligent system enough to distinguish people from other objects and people. It is able to identify the target person in the case of several people in a dynamic and capable of driving environment itself behind the targeted person carefully in order to follow him or her throughout by maintaining a safe distance. Any person following robotic system essentially has two major tasks person tracking or detection.

1.1 Person-Following.

The two modules that perform these tasks work in harmony to create a single independent system that recognizes its target, understands the commands and correctly follows the target by self-

Corresponding Author: M. S. Zaghloul, Electronics and Communication Department, Arab Academy for Science, Technology, and Maritime Transport, P.O.1029, Alex, Egypt.
E-mail: dr_mszaghloul@yahoo.com

adjusting its motion based on the movement of the target. Person-tracking can further be separate into person detection and location-tracking. There are several challenges in designing a person tracking system. For example, a person following robot should be functioning properly in a crowded environment. the person's face, which, as a matter of fact, is the most unique feature of a human body It should not be lost during tracking when there is a lot of movement in front. Furthermore, person following robot designer should consider a place where there is uneven floor. A person following robot should also be able to increase its speed if the person is running. The robot should also avoid obstacles, be it static or moving obstacles. Sometimes, there will be a person who is much similar to the target person, and may be wearing similar clothes. The robot should be able to distinguish and follow the correct person. The space of the robot and the person followed should be in a safe distance. The robot should know when to stop, and avoid collision with the person. As a result, it limits the amount of data available to perform the recognition. The only visual clue obtainable is the image of the person from the rear and hence the person's outfit becomes the single choice to accomplish accurate person-detection.

2-Previous work and Current Trend

Any person following robotic system essentially has two major tasks –person tracking or detection and person-following. The two modules that perform these tasks work in unison to create a single autonomous system that recognizes its target, understands the commands and correctly follows the target by self-adjusting its motion based on the movement of the target. Person-tracking can further be segregated into person detection and location-tracking. Person detection involves distinction of humans from other objects and it may also include recognition of a particular person in a group of people, in case of tracking in crowds. The person following part involves defining the adaptive behavior of the robot motion in real time i.e. how a robot would follow a particular person safely in a dynamic and non-predetermined environment. There are several challenges in designing a person tracking system. First of all, a robot can never see the person's face, which, as a matter of fact, is the most unique feature of a human body. As a result, it limits the amount of data available to perform the recognition. He only visual clue obtainable is the image of the person from the rear and hence the person's outfit becomes the single choice to accomplish accurate person-detection. This also means that a recognition module based solely on the outfit matching technique is bound to fail when people are in uniform. It also requires the color of the outfit to be different from the 2 color of the background color which may not always be possible. Secondly, the change in light conditions has a significant effect on the color data which becomes a major concern in case of person tracking and following system because the target can move around in different light settings. Although there are a few known techniques, like the use of HSI color model instead of RGB, to minimize the effect of change in background lighting on the color information, these methods don't guarantee color data to be completely light-invariant. Thirdly, for shape-based tracking (that only works in case of a single person), the system must know about all the shapes and positions of a human body in order to correctly identify it as a person. This kind of learning requires tremendous amount of data which is not easy to gather (Microsoft Kinect's person identification is based on this technique)(Maccormick, 2013; Shotton et al., 2011). A simpler alternative is to only identify the shape of the head and then track that shape but it may lead to false detections when there is an oval or round- shaped object in the background (Shenoy, 2013). The person-following task comes with its own set of challenges. The most important responsibility of the robot is to shadow the person by maintaining a safe distance from him/her to avoid collisions. In real-world situations, it is not easy to do so because the robot has to deal with several factors like unstructured spaces, unknown external conditions like terrain and obstacles and changing walking speeds of the target. The robot needs to continuously adapt to the human walking speed in such a way that it is able to go fast enough to not get left behind and slow enough to not get past the target.

3-Proposed System

Block Diagram

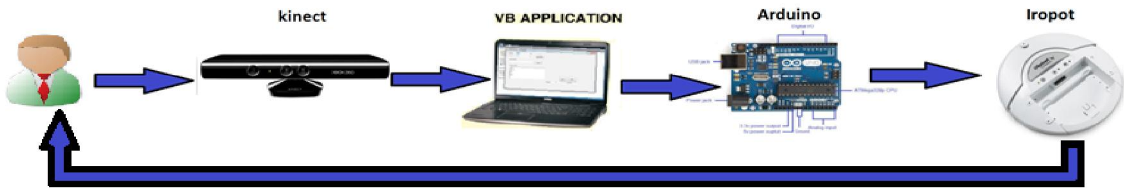

Fig. 1: System Block Diagram

3.1. System Hardware

Microsoft Kinect (XBOX 360):

The infrared and visible spectrum cameras on Kinect, along with its firmware allow it to distinguish humans from other objects in scene Kinect hardware (Eltoukhy *et al.,* 2016) key features (Kinect, 2017). Additionally, its infrared sensor allows it to determine the distance between it and the objects in the scene. Moreover, the Kinect Software Development Kit (SDK) (Kinect, 2017) released by Microsoft to help developers write software for it was also used extensively during the course of the development of our robot.

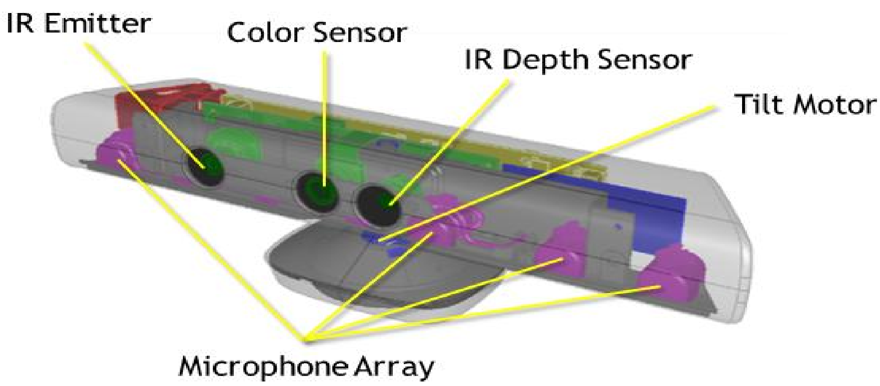

Fig. 2: Represents Microsoft Kinect

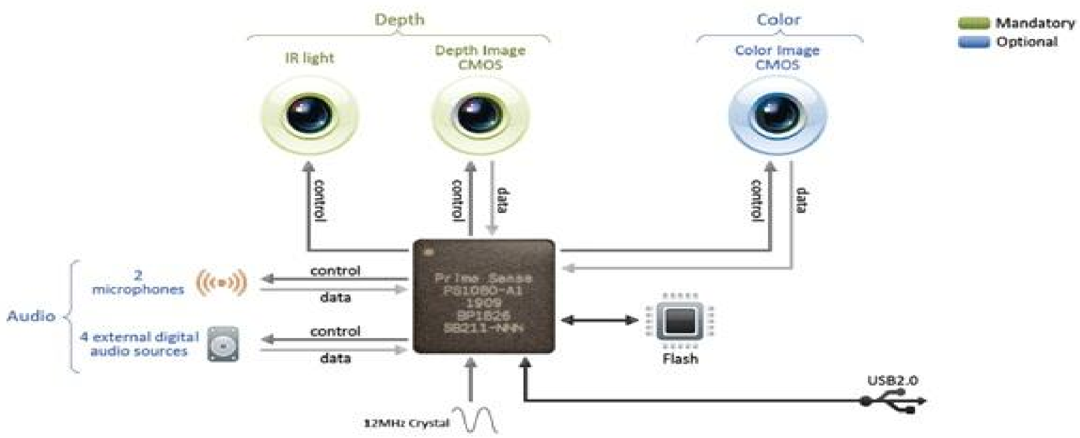

Fig. 3: Represents Microsoft Kinect Block Diagram

Arduino programmable microcontroller:

Board as shown in figure 3: This board package provides a microcontroller along with a number of digital input and output pins to help interface with other electronic components (Arduino, 2017; Nayyar *et al.,* 2016). In this project, it was used to build a light intensity sensor whose readings were used to provide data pertaining to illumination that was used to enhance the accuracy of the image recognition and comparison algorithm (Chaudhary and Sarwar, 2013).

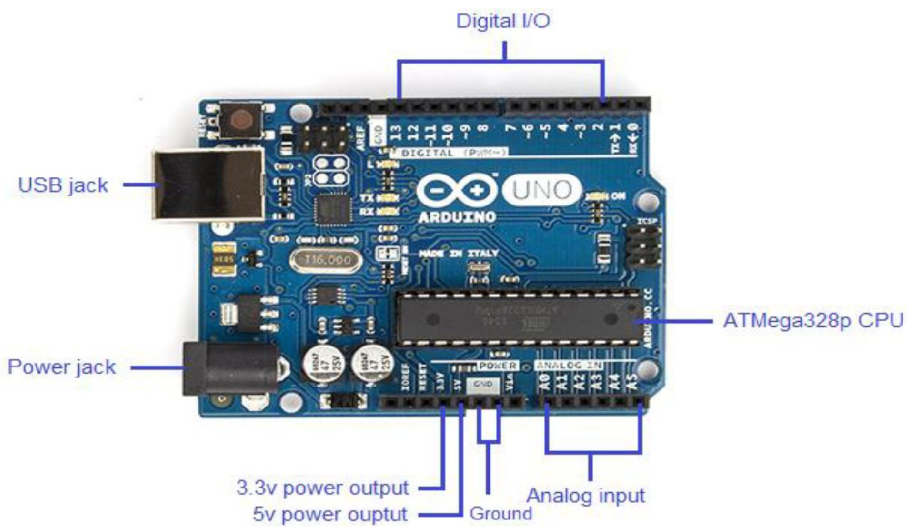

Fig. 4: The Arduino ATMega microcontroller

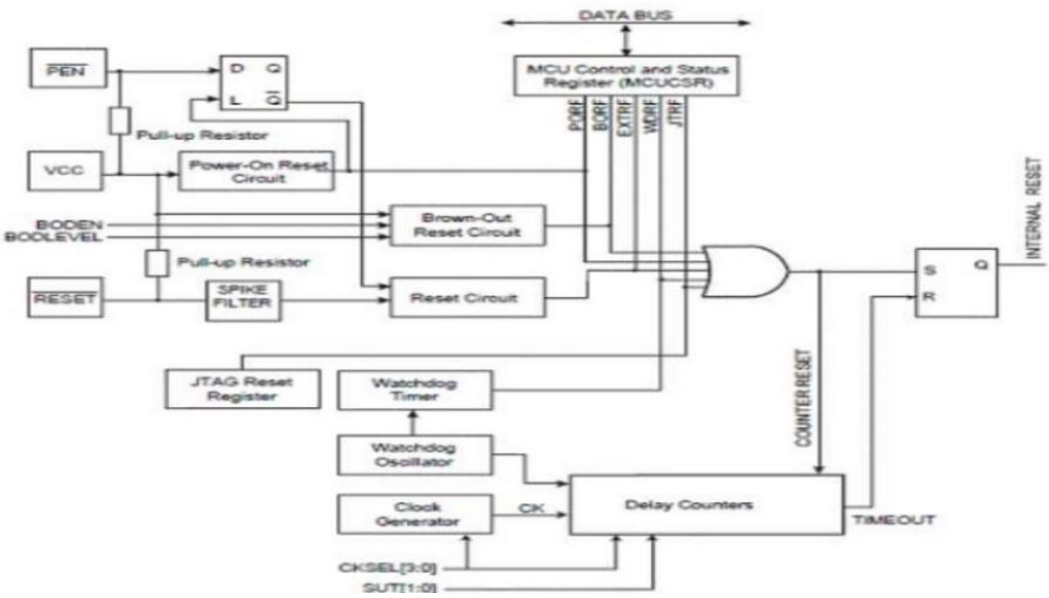

Fig. 5: Arduino Block diagram

IRobot Create programmable robot:

As shown in figure 4: While the sensing and recognition algorithms provided the ability to identify which human in the scene to follow, the actual movement was carried out by the iRobot

Create robotic vehicle. It most important features were the differential motors that allowed the human-following robot to rotate about its axis and the serial interface that could be used to communicate with the computer that was doing the recognition (Iropot, 2017).

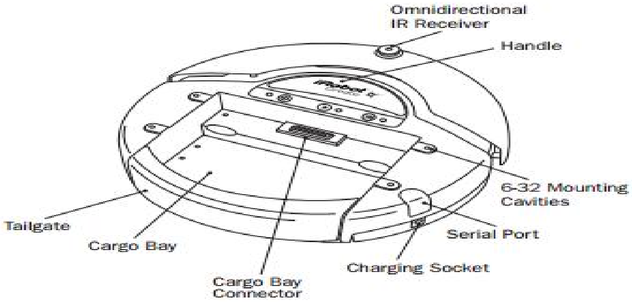

Fig. 6: IRobot configuration

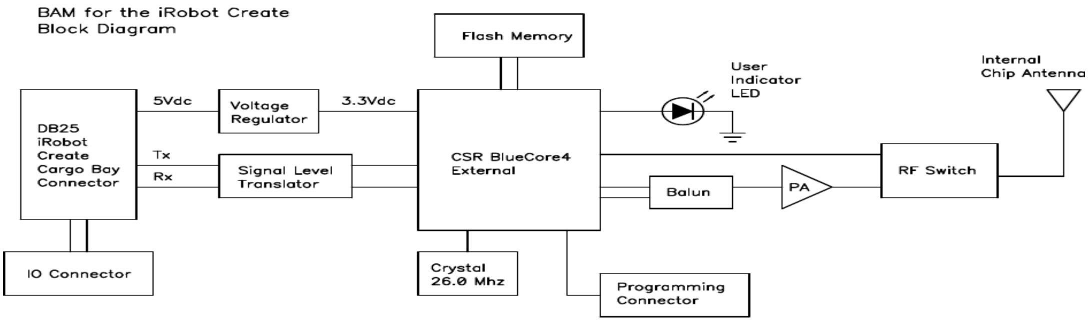

Fig. 7: IRobot Block Diagram

Data Flow

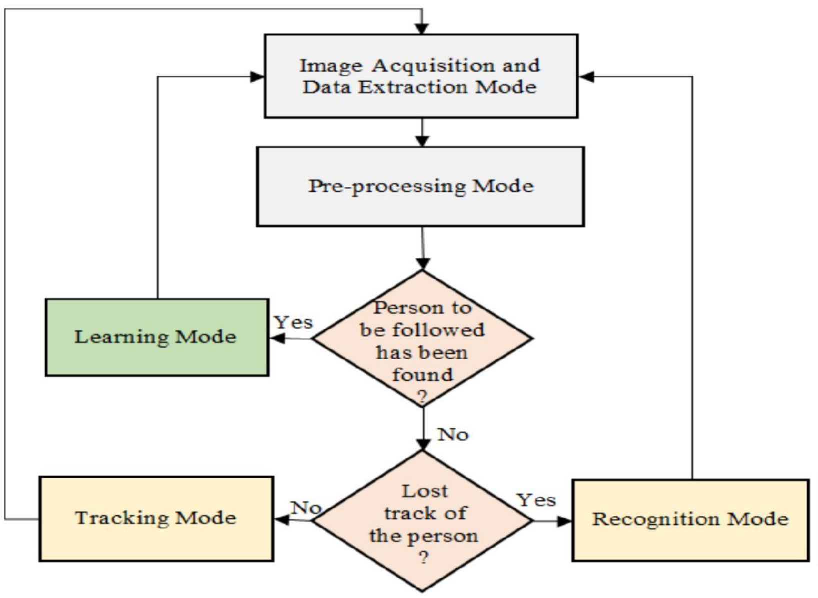

Fig. 8: Person-tracking algorithm

3.2 System Software

Person-tracking and following system is one that uses data collected from various sensors to identify someone in a group and translates their motion information to guide the motion of the robot.

Most vision-based person-following systems acquire data using lenses and sensors, identify people based on features or gestures, track them in changing background conditions and command a robot to follow that person. The person-tracking algorithm can be divided into the following stages or modes: Image-Acquisition and Data-Extraction mode, Pre-processing mode, learning mode, tracking mode and Recognition mode. The first two stages - Image Acquisition and Data Extraction and Pre-processing, are executed sequentially for every new frame of image data. The algorithm then chooses which mode to enter from one of the remaining three modes based on the current processing step in the program as shown in figure 5.

- The image acquisition hardware used is Microsoft‘s Kinect system. It provides a 640 x 480 pixel RGB Image, a 640 x 480 infrared image and a raw depth map. The resolution of images is inversely dependent on the frame rate i.e. for a lower frame rate, the resolution is higher. In this project the frame rate is set to 30 frames per second for both depth and color streams (Ghanbarnezhad et al., 2016).
-Pre Processing Mode The pre-processing mode consists of steps that select, filter and format the color and skeleton stream data, in order to extract relevant information from it. It is divided into the following processing steps - Background Subtraction, Torso Area Detection or Region of Interest (ROI) Extraction, Re-Quantization, RGB to HSI Conversion and Color and Skeleton Image Construction which are performed in a sequence on every new image frame
-Tracking mode once the person to be followed is located and all the data about that person is recorded, the program enters the tracking mode where it performs two tasks – track the person and update the histograms when the light condition changes (Microsoft Developer Network, 2017).
-Recognition Mode During the tracking phase, if the image of the person being tracked gets lost (which is indicated by an audio feedback), then the program enters recognition mode where it first sends a command to the robot to stop its movement and then tries to locate him/her in the new frames. This means that every incoming frame of data needs to be analyzed for a match with the visual characteristics of the individual marked to be followed in the learning phase (Zhou and Fan, 2016).
-Learning mode where it tries to locate the person of interest in every new image frame and if he/she is detected then it tries to learn about the visual characteristics of that person. The algorithm performs pose-based recognition in order to locate the person to be followed

4- Results

We were able to implements a person-tracking and following robotic system using vision based and sensor-based techniques. It demonstrates several contributions. First, it combines all the heterogeneous forms of data provided by various sensors so as to gather information about people in the frame and to recognize the target person. More specifically, it gathers the outfit color histogram data from the color-sensor and the shape, size and location of the human skeleton from the infrared-sensor in order to track and locate the target. Second, the use of different sensors, on the one hand, augments some of the existing recognition techniques like color-based histogram matching, and on the other hand, it provides the opportunity to use new ideas that can leverage the data provided by other sensors. Third, the system can learn about the target without any manual intervention. It is capable of recognizing poses and hence it can start tracking anyone who is detected to be in a tracking pose (which is currently set as the frontal or posterior pose with both hands up at a 90-degree angle). Finally, it is designed to be a self-navigating system that can drive itself behind the target by maintaining a safe distance from him or her. It maps the changes in human walking speed to appropriate commands for adjusting the robot‘s movements to be able to follow the target person around. It takes into account the erratic motion changes of the target and instead of reciprocating with a similar jerky movement; it tries to respond in a graceful way using gradual step-up and step-down mechanism.

As shown in figures 5 to 9 which represent the different combinations for tracking person at different distance, angles or views and approaching or departing. The tracking was successful for all the demonstrated combinations.

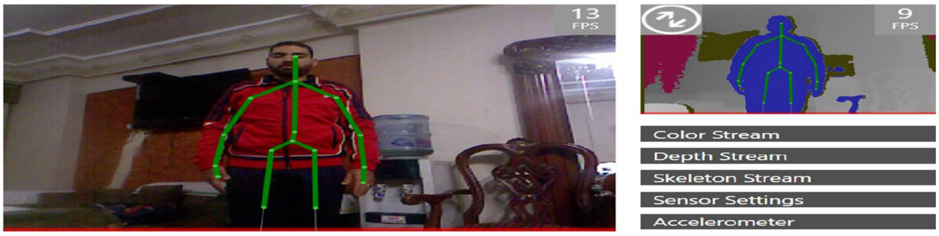

Fig. 9: Represent Microsoft kinect recognition when a person about 0.5 meter distance

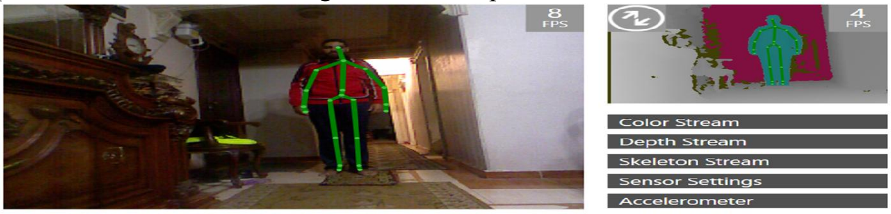

Fig.10: Represents Microsoft kinec recognition when a person at around 4 meters distance

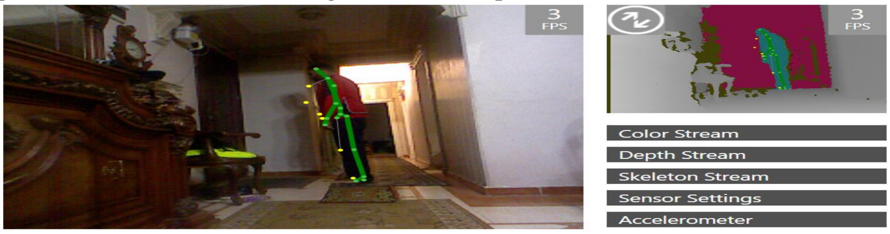

Fig. 11: The Microsoft kinec recognition and tracking when looking to a person from the side view

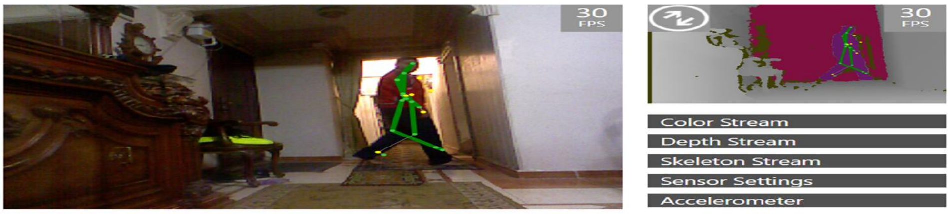

Fig. 12: Microsoft kinect the tracking person when tracked person try to escape with side view from iRobot

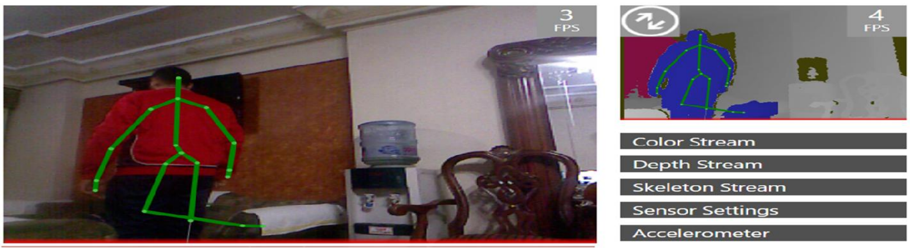

Fig. 13: Microsoft kinect tracking person when a tracked person try to escape from IRobot but IRobot flow him

Change in Height with Change in Depth

In order to make accurate height comparisons, the height value should be independent of the distance from the camera lens because the size of the person appears to be decreasing with increasing depth from the camera lens. This is in accordance with the Inverse-square law [41] that is stated as,

$$Apparent\ size = \frac{Actual\ Size}{d^2}$$

Where d is the distance from the camera lens. The acceptable d from 6ft to 12 feet.

Effect of Change in Light Conditions

The biggest challenge in any computer vision and image processing based application is to minimize the effect of change in background lighting on the color data. Variations in light conditions have significant impact on color information. For example, in low lighting, the dark shades may appear all black. The test is set up in a room by creating a controlled lighting environment using floor lamps with dimmers. A Nexus 5 phone measures the LUX (SI unit of luminance) value as the light conditions change. Assessment begins by setting the light intensity of all the floor lamps to the maximum, resulting in a room luminance of LUX value 5.

Tracking Pose Detection Accuracy Test

The algorithm performs tracking pose evaluation at the beginning of the person following process. It is needed to determine and locate the target person who needs to be tracked and followed by the robot. Detection of a tracking pose is nothing but the calculation of the angle between the upper arm and the lower arm at the elbow joint when the hands are up in the air. It is said to be a tracking pose only when this angle is 90 degrees (with a 20- degree tolerance either way)as shown in figure 11. To perform angle calculation, the joint coordinate information of left and right shoulders, left and right elbows and left and right wrists is required. Figure 5.6 shows the poses that are considered to be valid for detecting and tracking the target person.

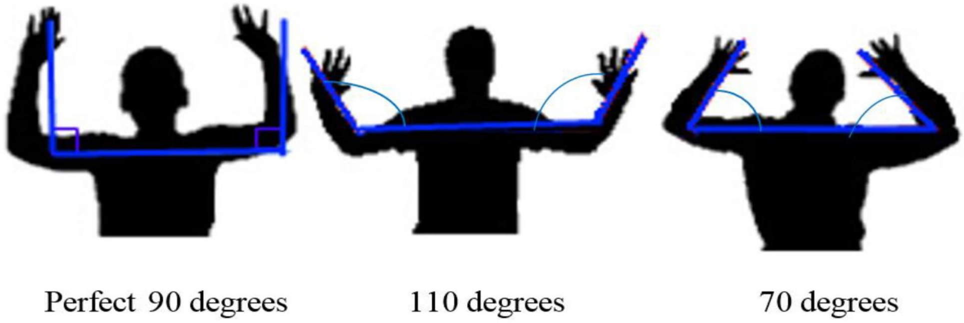

Fig. 14: Tracking pose angle

Table 1: Obtained tracking result and comparison with previous results

Experiment Number	Number of person	Test condition	Number of times repeated	Number of times passed	Other previous algorithm (Lnu,2014) Number of times repeated	Other previous algorithm (Lnu,2014) Number of times passed	Comment
1	1	Constant lighting	5	5	2	2	Both results are equal 100%
2	1	Constant lighting but the person was standing next to a big object	10	8	5	1	Our result represent 80% and pervious result represent 20%
3	1	Variable lighting between 4 lux to 6 lux	10	7	-	-	Our result is 70% while this test was not applied for the previous work
4	1	In the sunlight	10	0	2	0	Both result are equal to zero Because No skeleton image was produced
5	2	Constant lighting	2	2	3	2	Our result represent 100% and pervious result represent 66.6%
6	2	Constant lighting but the person was standing next to a big object	4	3	-	-	Our result is 75% while this test was not applied for the previous work
7	2	Variable lighting between 4 lux to 6 lux	5	4	2	1	Our result represent 80% and pervious result represent 50%
8	2	In the sunlight	3	0	2	0	Both result are equal to zero Because No skeleton image was produced

9	**1**	Constant walking speed in open space	3	3	2	2	Both results are equal 100%
10	**1**	Sudden speed changes	3	0	2	0	Both result are equal to zero because Lost tracking (the person had to come back in the frame to get recognized again)
11	**2**	Constant speed, walking very close to each other	4	1	2	0	Our result represent 25% and pervious result represent 0% because Kinect was able to track with low percentage for one person due to being very close
12	**1**	Constant speed, moved from carpet to hardwood floor	3	2	2	2	Our result represent 66.6% and pervious result represent 100% Because our system is heavy and has difficulty to move on a hard wood
13	**3**	Constant speed, moved from carpet to hardwood floor	4	3	2	2	Our result represent 75% and pervious result represent 100% Because our system is heavy and has difficulty to move on a hard wood
		Sum	66	38	26	12	Overall our results success represent 57.57% previous results success represent 46.15% so improvement ratio with 24.74%

As shown in table 1 the number of common trails are 11 and we apply extra 2 trails was not done in the previous work The percentage of success is shown in figure 12

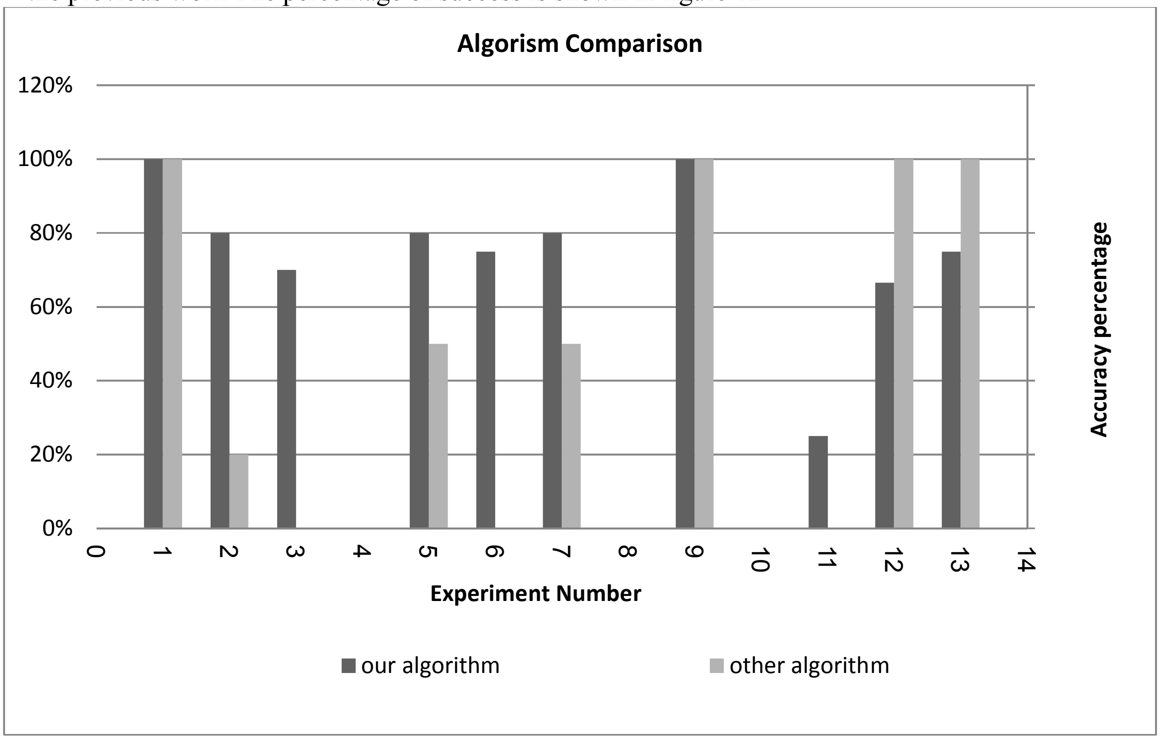

Fig. 15: Algorism Comparison

5- Conclusion

The main contribution of the paper is the proposal of novel ideas to perform a more reliable recognition on the basis of the color and skeleton images of the people present in the scene. It records changes in light intensity by monitoring the readings of a light dependent resistor and the average intensity of the torso area and then uses this information to perform a light invariant hue and saturation histogram matching. The use of an infrared sensor gives enough information to extract a skeleton image of the person and this in turn can be used to perform body-shape and size comparisons of the people in the frame against the target person‘s skeleton. Adding to that the system was designed to be a self-navigating system that can drive itself behind the target by maintaining a safe distance from him. Moreover the changes in human walking speed can be mapped to appropriate commands for adjusting the robot's movements to be able to follow the target person around. We can conclude that our work with different trails has better results as shown in table 1 and graph number AAAA compared with the previous work. Moreover we had done more trails as Variable lighting between 4 lux to 6 lux and Constant lighting and when the person was standing next to a big object, for theses to cases we got success from 70% to 75%

References

Arduino, 2017. The Arduino Mega is a microcontroller board based on the ATmega128 https://www.arduino.cc/en/Main/arduino Board Mega, accessed July.

Chaudhary, A. and S. Sarwar, 2013. A Sensor Based Usb Driven Lamp Using Computer Hardware Technology'. International Journal of Engineering Research & Technology, May.

Eltoukhy, M. and A. Kelly, C.-Y. Kim, H.-Pil Jun, R. Campbell and C. Kuenze, 2016. Validation of the Microsoft Kinect camera system for measurement of lower extremity jump landing and squatting kinematics. Journal Sports Biomechanics , 02 Feb.

Ghanbarnezhad, F., S. Hosseini and M. Hosseini, 2016. Recognition of Abnormal Human Behavior using Kinect Case Study: Tehran Metro Station. Indian Journal of Science and Technology, November.

Iropot, 2017. irobot Creat Manual, http:// www. irobot. com/f ilelibrary/ create /Create %20 Manual_Final.pdf, accessed July.

Kinect, 2017. Kinect hardware key features. https://developer.microsoft.com/en-us/ windows /kinect/ hardware, accessed July.

Kinect, 2017. The Kinect Software Development Kit (SDK) https:// developer. Microsoft. com/en-us/windows/kinect, accessed July .

Lnu, V., 2014. Robotic Person Following in a Crowd Using Infrared and Vision Sensors' . Master of Science in Computer Science, San Diego State University.

Maccormick, J., 2013. How does the Kinect work?. University of Wisconsin, Madison, December.

Microsoft Developer Network, 2017. Tracking users with Kinect skeleton tracking. Microsoft, http://msdn.microsoft.com/en-us/library/jj131025.aspx, accessed July.

Nayyar, A. and V.Puri, 2016. A review of Arduino boards, Lilypad's & Arduino shields. Computing for Sustainable Global Development (INDIACom), 2016

Shenoy, R., 2013. Enhancing the autonomous robotic person detection and following using modified Hough transforms '. Master's thesis, San Diego State University, San Diego, CA.

Shotton, J., A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake, 2011. Real-time human pose recognition in parts from single depth images. Microsoft , Washington .

Yejin, K., S. Baek, B.-Chull Bae, 2017. Motion Capture of the Human Body Using Multiple Depth Sensors. ETRI Journal, Number 2, April.

Zhou, Y. and C. Fan, 2016. Action recognition robust to position changes using skeleton information and SVM. Robotics and Automation Engineering (ICRAE), 27-29 Aug.