Article

ROS-Based Human Detection and Tracking from a Wireless Controlled Mobile Robot Using Kinect

Sidagam Sankar [1,†] and Chi-Yi Tsai [2,*,†]

[1] Department of Electronics and Communication Engineering, Vel tech Rangarajan Dr. Sagunthala R&D Institution of Science and Technology, No.42, Veltech Road, Avadi, Chennai 600062, Tamil Nadu, India; sidagamsankar@gmail.com

[2] Department of Electrical and Computer Engineering, Tamkang University, No.151, Yingzhuan Road, Tamsui District, New Taipei City 25137, Taiwan

[*] Correspondence: chiyi_tsai@mail.tku.edu.tw

[†] The authors contributed equally to this work.

check for updates

Abstract: Human detection and tracking is an important task in artificial intelligent robotic systems, which usually require a robust target detector to work in a variety of circumstances. In complex environments where a camera is installed on a wheeled mobile robot, this task becomes much more difficult. In this paper, we present a real-time remote-control system for human detection, tracking, security, and verification in such challenging environments. Such a system is useful for security monitoring, data collection, and experimental purpose. In the proposed system, a Kinect RGB-D camera from Microsoft is used as a visual sensing device for the design of human detection and tracking. We also implemented a remote-control system on a four-wheel mobile platform with a Robot Operating System (ROS). By combining these two designs, a wireless controlled mobile platform having the feature of real-time human monitoring can be realized to handle this task efficiently. Experimental results validate the performance of the proposed system for wirelessly controlling the mobile robot to track humans in a real-world environment.

Keywords: robot operating system (ROS); human detection; human tracking; wheeled mobile platform; RGB-D camera

1. Introduction

Automated surveillance for security monitoring will increase more in venues such as airports, casinos, museums, and public places, etc. In these places, the government usually requires a surveillance system installed with intelligent software to monitor security and to detect suspicious behavior of people using cameras. In a remote-control center, human operators could search archived videos to find abnormal activities of people in public places such as college campuses, transport stations, and parking lots, etc. Having this type of surveillance system vastly increases the productivity of the human operator and expands coverage of the security monitoring service. However, the problem of human detection and tracking in video sequences becomes much harder when the camera is mounted on a mobile platform moving in a crowded workspace. If the object of interest is a deformable object, like a human, the problem becomes even more challenging to detect and to track the human.

A modern mobile robot is usually prepared with an onboard digital camera and computer vision software for human detection and tracking, which can be applied in several real-world applications. In this paper, a real-time computer vision system for human detection and tracking on a wheeled mobile platform, such as a car-like robot, is presented. The contribution presented in this paper is mainly focused on the system design and implementation aspects. In other words, our design aims to

achieve two main goals: Human detection and tracking, and wireless remote control of the mobile platform. Human detection and tracking are achieved through the integration of point cloud-based human detection, position-based human tracking, and motion analysis in the robot operating system (ROS) framework [1]. On the other hand, the wireless remote control of the mobile platform is achieved by the HyperText Transfer Protocol (HTTP) [2]. Developing such an intelligent wireless motion platform is very useful for many practical applications such as remote monitoring, data collection, and experimental purpose, etc. Experimental results validate the performance of the proposed system for detecting and tracking of multiple humans in an indoor environment.

The rest of this paper is organized as follows. Section 2 explores the literature review on some of the related literature to this study. Section 3 explains our system design in detail. In Section 4, we explain details about the implementation of human detection, human tracking, and remote control of the mobile platform. Section 5 presents the experimental results of the proposed system with some discussions. Section 6 concludes the contribution of this paper followed by the future scope of this study.

2. Literature Survey

In this section, we discuss the literature survey related to the design of human detection and tracking systems. Mohamed Hussein et al. [3] proposed a human detection module that is responsible for invoking the detection algorithm. Ideally, the detection algorithm is to be run on each input frame. However, this will inhibit the system from meeting its real-time requirements. Instead, the detection algorithm in their implementation is invoked every two seconds. The location of the human targets at the remaining time is determined by tracking the detected humans using the tracking algorithm. To further speed up the process, the detection algorithm does not look for humans in the entire frame. Instead, it looks for humans in the regions determined to be foreground regions. The tracking module processes frames and detections received from the human detection module to retain information about all the existing tracks. When a new frame is received, the current tracks are extended by locating the new bounding boxes for each track location in the new frame. If the received frame is accompanied with new detections, the new detections are compared to the current tracks. If the new detection significantly overlaps with one of the existing tracks, it is ignored. Otherwise, the new track is created as a new detection. A track is terminated if the tracking algorithm fails to extend it in a new input frame.

Juyeon Lee et al. [4] proposed a human detection and tracking system consisting of seven modules: Data collector (DC) module, shape detector (SD) module, location decision (LD) module, shape standardization (SS) module, feature collector (FC) module, motion decision (MD) module, and a location recognition (LR) module. The DC module acquires color images from the digital network cameras every three seconds. The SD module analyzes the human's volume from four images, which is acquired from the DC module. They used a moving window to extract the human's coordinates in each image. The LD module analyses the user's location through coordinates that are offered by the SD module. The LD module also analyses whether the human is near to which home appliance through a comparison between the absolute coordinate of the human and the absolute coordinate of the appliance in the home. The SS module converts the image into a standard image for recognition of multi-users' motion. The FC module extracts the image that is changed by the SS module, and the MD module predicts the motion of the detected human. The SD module receives raw images from the DC module to distinguish the human's image through subtraction between the second image and the third image, then calculates the human's absolute coordinates in the home and the human's relative coordinates with reference to the furniture and home appliances. The SD module also decides which furniture the human is near to. If there is a human in an important place that is defined for the human tracking agent, the SD module calculates relative coordinates of the moving window, which includes a human, and transmits those to the LD module. The LR module can judge a user's location without conversion of the acquired image. Unlike the human's location, the human's motion should

be recognized to differ in the case of each person because a person's appearance differs from each other. The SS module takes charge of motion recognition in the multi-user situation. To recognize the human's motion, they defined feature sets of six standardized motion showing six motions that are recognized in the human tracking.

Ramesh K. Kulkarni et al. [5] proposed a human tracking system based on video cameras. There are three steps in the proposed system. The first step is a Gaussian-based background subtraction process [6] to remove the background of the input image. The second step is the detection of a targeted human using a matching process between the foreground and the target image based on the Principle Component Analysis (PCA) algorithm [7], which is a method of transforming a number of correlated variables into a smaller number of uncorrelated variables based on eigen-decomposition. The final step is using a Kalman filtering algorithm [8] to track the detected human.

In reference [9], Sales et al. proposed a real-time robotic system capable of mapping indoor, cluttered environments, detecting people, and localizing them in the map using a RGB-D camera mounted on top of a mobile robot. The proposed system integrates a grid-based simultaneous localization and mapping approach with a point cloud-based people detection method to localize both the robot and the detected people in the same map coordinate system.

In reference [10], Jafari et al. proposed a real-time RGB-D based multi-person detection and tracking system suitable for mobile robots and head-worn cameras. The proposed vision system combines two different detectors for different distance ranges. For the close range, the system enables a fast depth-based upper-body detector, which allows real-time performance on a single core processing unit (CPU) core. For the farther distance ranges, the system switches to an appearance-based full-body Histogram-of-Oriented-Gradient (HOG) detector running on the graphic processing unit (GPU). By combining these two detectors, the robustness of the vision system using an RGB-D camera can be improved for indoor settings.

Revathy Nair M et al. [11] proposed a stereo vision based human detection and tracking system, which is integrated with the help of image processing. To identify the presence of human beings, an HSV-based skin color detection algorithm is used. When the system identifies the presence of human beings, the human location is then identified by applying a stereo vision-based depth estimation approach. If the intruder is moving, the centroid position of the human with the orientation angle of motion is obtained. In the tracking part, the human tracking system is a separate hardware system connected to two servo motors for targeting the intruder. Both human detection and human tracking systems are separately coupled through a serial port. If there are several people present on the video or image, then the centroid of all the regions are estimated and the midpoint of them gives the required result.

In reference [12], Priyandoko proposed an autonomous human following robot, which comprises face detection, leg detection, color detection, and person blob detection methods based on RGB-D and laser sensors. The authors used Robot Operating System (ROS) to perform the four detection methods, and the result showed that the robot could track and follow the target person based on the person's movement.

As to the wireless control of the mobile platforms, Shayban Nasif et al. [13] proposed a smart wheelchair, which can be useful assistance to physically disabled persons to complete their daily needs. The previous development of this smart wheelchair system was developed using a personal computer or laptop in the wheelchair. Because recent developments in embedded systems can be combined to the design of robotics, artificial intelligence, and automation control, the authors further integrated a Radio Frequency (RF) communication module into the smart wheelchair system to implement wireless control function in the wheelchair. Moreover, the authors used an acceleration sensor to implement a head gesture recognition module, which allows the user to control the wheelchair in directions as needed to move, using the change of head gesture only.

In reference [14], Patil et al. proposed the hardware architecture for a mobile heterogeneous robot swarm system, in which each robot has a distinct type of hardware compared to the other robots.

One of the most important design goals for the heterogeneous robot swarm system is providing low cost wireless communication for indoor as well as outdoor applications. To achieve this goal, the authors adopted X-Bee module, Bluetooth Bee module, and PmodWiFi module to create an ad hoc communication network, which can efficiently accomplish the wireless communication for decentralized robot swarms.

3. System Design

Our design aims to achieve two main goals: Human detection and wireless remote control of a wheeled mobile platform. Human detection is achieved through the integration of 3D point cloud processes for human detection, tracking, and motion analysis, in one framework. Wireless remote control of the mobile platform is achieved by using an HTTP protocol.

Figure 1 shows the architecture of the proposed system design. We use a Kinect device to capture 640 × 480 24-bit full-color and 320 × 240 16-bit depth images, which are used as the input of the proposed system. A laptop is used as a major image processing unit in this system to receive the captured visual sensing data through USB 2.0 port, which provides a high frame rate up to 30 frames per second (fps) for transmitting both full-color and depth images. The function of human detection and tracking is implemented in the laptop to detect the live motion of the humans in front of the mobile platform. This function helps the proposed system for using in real-time applications. The output result of live human detection and tracking can be observed on the display device connected to the laptop.
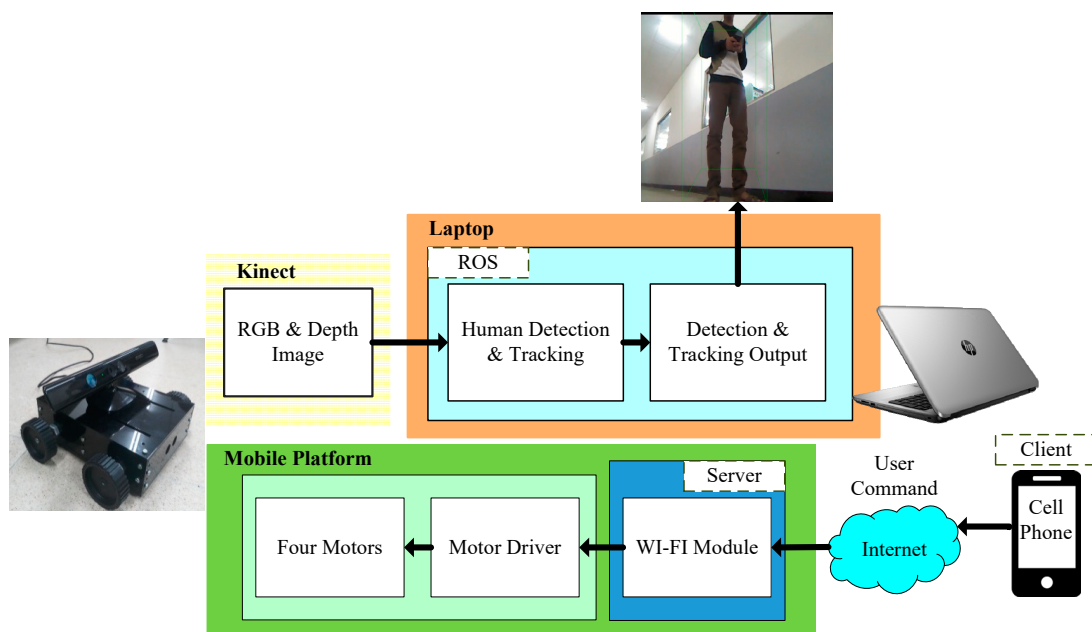

Figure 1. System Architecture.

As for the mobile platform, the Nodemcu esp8266 WI-FI module plays a major role. This module is integrated with integrated circuit (IC) and WI-FI shield on it. The IP address of the WI-FI module must be browsed in the cell phone, and then we will get the page with a control interface. There are four motors used in the mobile platform. Each motor receives the driving signal from the motor driver, which is connected to the WI-FI module to help the user in controlling the mobile platform wirelessly. Note that the WI-FI module of the mobile platform should connect to the same network with the cell phone.

4. Implementation of Human Detecting and Wireless Mobile Platform Control

The human detection and tracking function used in the proposed system is implemented based on the ROS, which is a robotics middleware widely used in the robotics platform. It is a collection of

software functions for the development of robot software framework. The ROS is not an operating system; it provides services specially designed for a heterogeneous computer cluster such as the implementation of commonly used functionality, package management, hardware abstract and message-passing between processes. There are twelve ROS distributions released online, in which we installed the tenth ROS distribution named as Kinetic Kame in Ubuntu 16.04 Operating system.

4.1. Human Detection and Tracking

Human detection in a heavy traffic environment is complicated. When this task comes on a mobile platform, it will increase its complexity. In the implementation of this human detection function, we use the Kinect Xbox 360 device to capture the 3D point cloud data of the scene in front of the mobile platform through the RGB and depth images. The Kinect device is interfaced by the Open Natural Interface (OpenNI), which is a type of middleware, including the launch files used for accessing the devices. OpenNI is compatible with several RGB-D cameras such as the Microsoft Kinect, Asus XtionPro, etc. These products are suitable for visualization and processing by creating a nodelet graph, which transforms the raw data into digital images and video outputs by using the device drivers.

Figure 2 shows the block diagram of the human detection process. In Figure 2, the point cloud data from the environment is first normalized based on the point cloud position in three coordinates by using the following formulas.

$$x_{diff} = x_{\max} - x_{\min}, \tag{1}$$

$$y_{diff} = y_{\max} - y_{\min,} \tag{2}$$

$$z_{diff} = z_{\max} - z_{\min}, \tag{3}$$

where $(x_{\min}, x_{\max})$, $(y_{\min}, y_{\max})$, and $(z_{\min}, z_{\max})$ denote the maximum and minimum point cloud position on the x-axis, y-axis, and z-axis, respectively. In this way, a normalization process is applied to point cloud data by

$$n_x = p_x/scale, \tag{4}$$

$$n_y = p_y/scale, \tag{5}$$

$$n_z = p_z/scale, \tag{6}$$

where (n_x, n_y, n_z) denotes the normalized point cloud data and (p_x, p_y, p_z) represents the input point cloud data on the x-axis, y-axis, and z-axis, respectively. The scaling factor is determined by the following formula

$$scale = \max(x_{diff}, y_{diff}, z_{diff})/2 \tag{7}$$

where $\max(x, y, z)$ returns the maximum value among x, y, and z. The normalized point cloud data are then used to form a normalized histogram through an orthogonal transformation constructed from the eigenvectors of the normalized point cloud data. Next, the PCA algorithm is applied to the normalized histogram to reduce the dimension of the histogram. Finally, a support vector machine (SVM) predictor [15] is applied on the low-dimensional normalized histogram to classify human and non-human classes with respect to the input point cloud data.
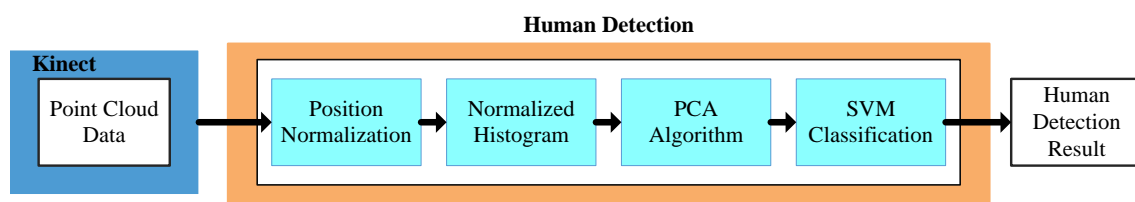

Figure 2. Block diagram of human detection process.

Figure 3 shows the block diagram of the human tracking process. As to the human tracking function, the output data of human detection are used as the input data to the nearest neighbor search (NNS) algorithm [16], which searches the nearest detected human position in the previous frame to the current detection result and then update the detected human position as the output tracking result. Those updated human positions are feedback as the previous human positions for the NNS algorithm to continue the human tracking process.
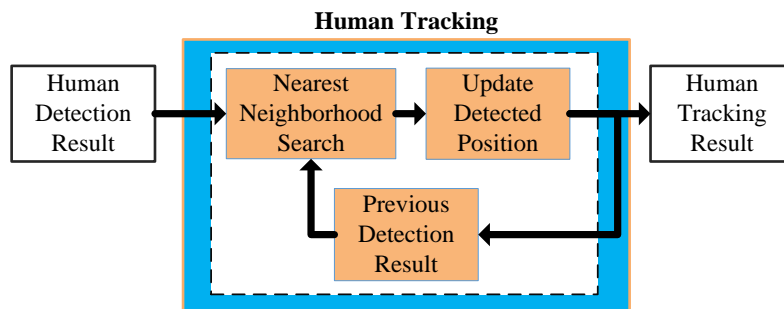
Figure 3. Block diagram of human tracking process.

4.2. Wireless Controlled Mobile Platform

The Nodemcu esp8266 microcontroller plays a crucial part in the implementation of the wheeled mobile platform, which uses four DC motors interfaced with an LM298N motor driver. The control signal of the motor direction given from the WI-FI module is converted into the transistor-transistor logic (TTL) by the motor driver for driving the motors in the required direction. Here, we use the Arduino IDE for programming the module.

To remote control the platform wirelessly, the cell phone and WI-FI module must connect to the same network. The IP address of the WI-FI module can be browsed in an HTML page. Figure 4 shows the HTML page with a control interface for the user to remotely control the mobile platform in the required direction wirelessly. The user is able to control the mobile platform wirelessly through a server-and-client based HTTP protocol, in which the cell phone acts as a client and the WI-FI module act as a server.
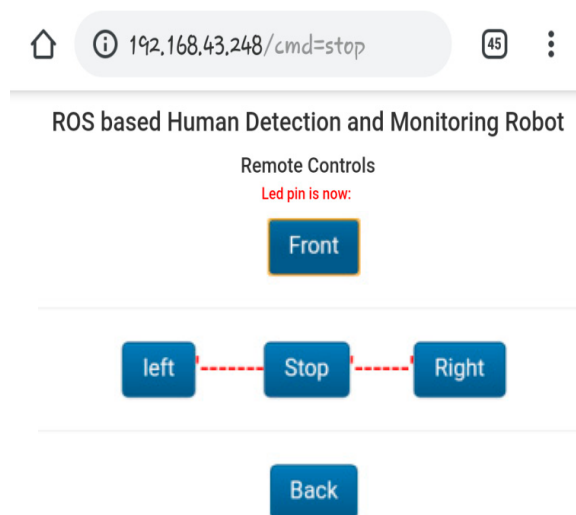

Figure 4. Remote control interface shown in the HTML page.

Figure 5 shows the flowchart of the remote control based on the control interface shown in Figure 4. The service set identifier (SSID) and Password are the details of the access point created by the user. Once the module is connected to the WI-FI network, it will check the server status and gives an IP address of the module, which is used in the client device to access the serve-and-client based

HTTP protocol. This workflow continuously reads the change of the user command. If the user inputs a new command from the control interface (Figure 5), then the controller will check all conditional statements. If the string is equal to the received command, then the controller starts to move the mobile platform until the next user command.
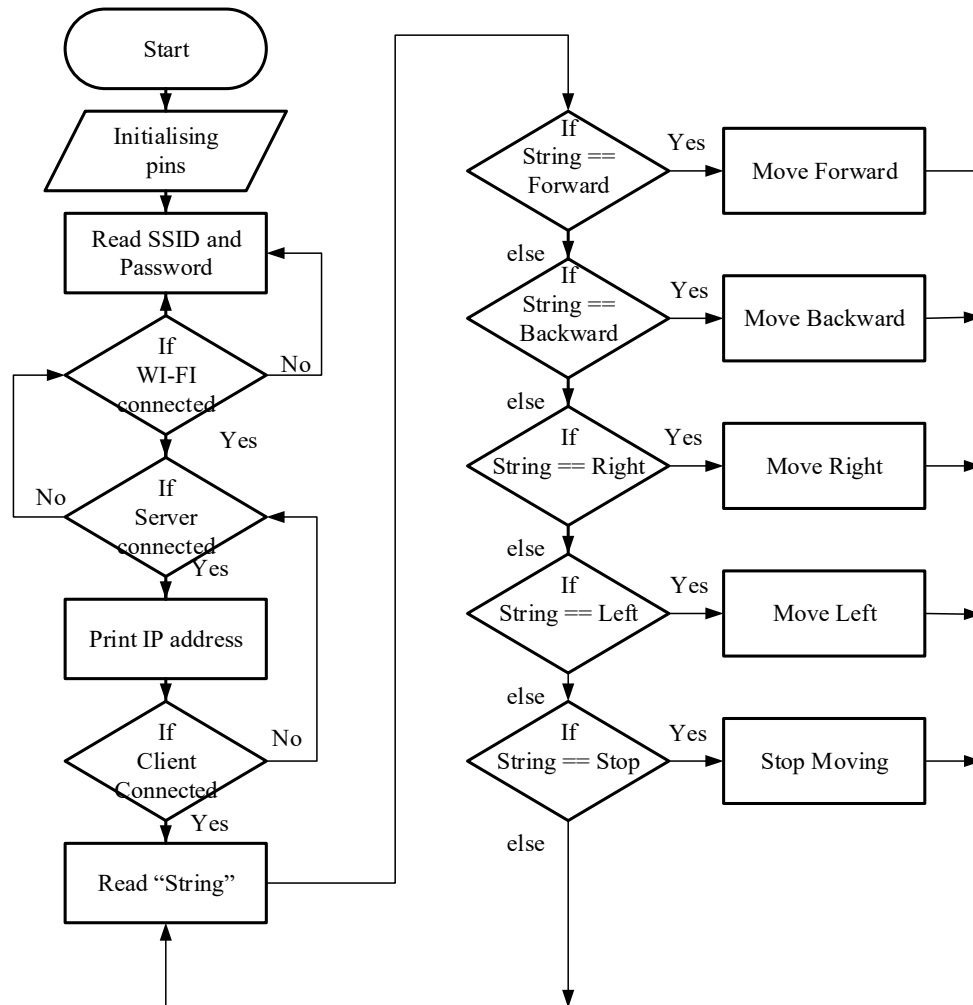
Figure 5. Remote control flowchart of the mobile platform.

5. Experimental Results

We implemented the ROS-based human detection and tracking system in a laptop equipped with a 2 GHz Intel®Core 64-bit i3-5005U CPU and 2 GB system memory running Ubuntu 16.04 operating system. The human detection and tracking result is displayed on the laptop screen through the ROS visualizer. Once the user launches the system, the ROS triggers all the algorithms to start detection and tracking the human. Two runtime scenarios are captured at two different locations in an indoor room to verify the performance of the proposed system. The Kinect device is mounted on the four-wheeled mobile platform to capture images for detecting and tracking humans in real time. Figure 6 shows the configuration of the four-wheeled mobile platform, which is a robotic vehicle chassis having four 12 volts DC motors produced by Nevonsolutions Pvt Ltd [17]. We further install a WI-FI module in this vehicle chassis to help in controlling the mobile platform wirelessly. Moreover, each DC motor is connected to an LM298N motor driver to help in driving the motors in the required direction according to the received control input through the WI-FI module.
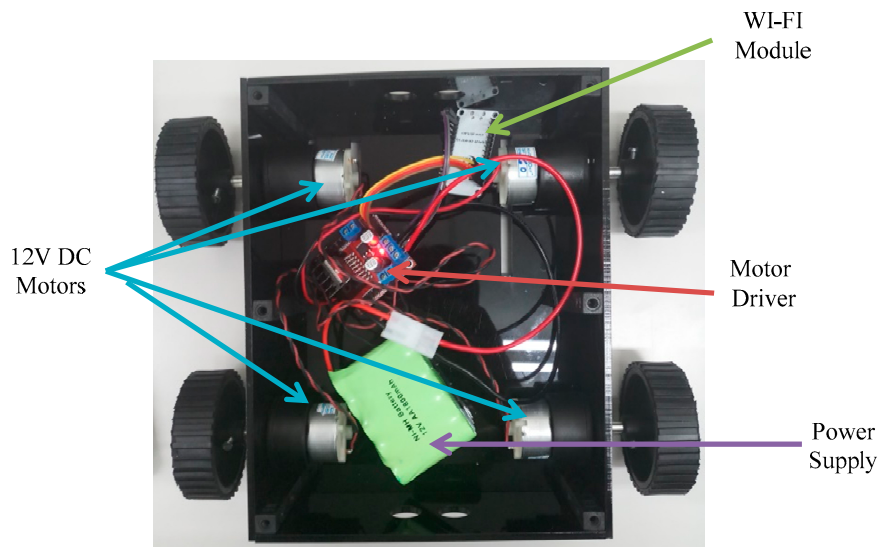
Figure 6. Configuration of the four-wheeled mobile platform.

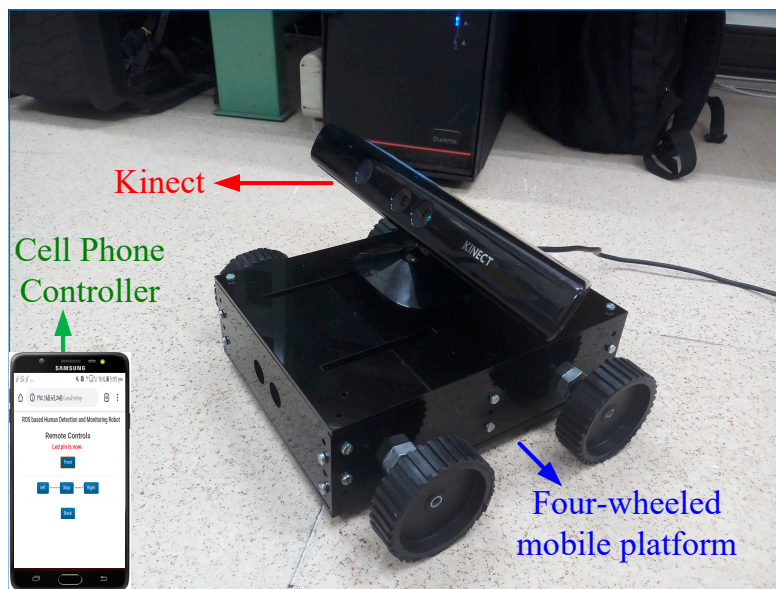Figure 7 shows the experimental setting used in the experiments. The four-wheeled mobile platform is equipped with a Kinect device and is controlled wirelessly by a cell phone with a control interface. The Kinect mounted on the mobile platform captures live video streams to detect and to track humans, whom can be monitored in the display of the laptop. The mobile platform is controlled through the cell phone via an HTML page for following the detected and tracked human.

Figure 7. Experimental setting: The four-wheeled mobile platform equipped with a Kinect device and controlled wirelessly by a cell phone with a control interface.

The first runtime scenario is an experiment of interaction between one detected human and the mobile robot through wireless remote control. Figure 8 shows the photos of motion sequences of interaction between the human and the remote-controlled mobile platform during this experiment. Figure 8a,c,e is the input images taken from the Kinect on the mobile platform. The sky-blue colored polygon box in Figure 8b is the human detection result of the proposed system, and the mobile robot started moving forward by the control command given through cell phone by a remote user. Figure 8d,f shows the mobile platform is in motion for tracking the human.
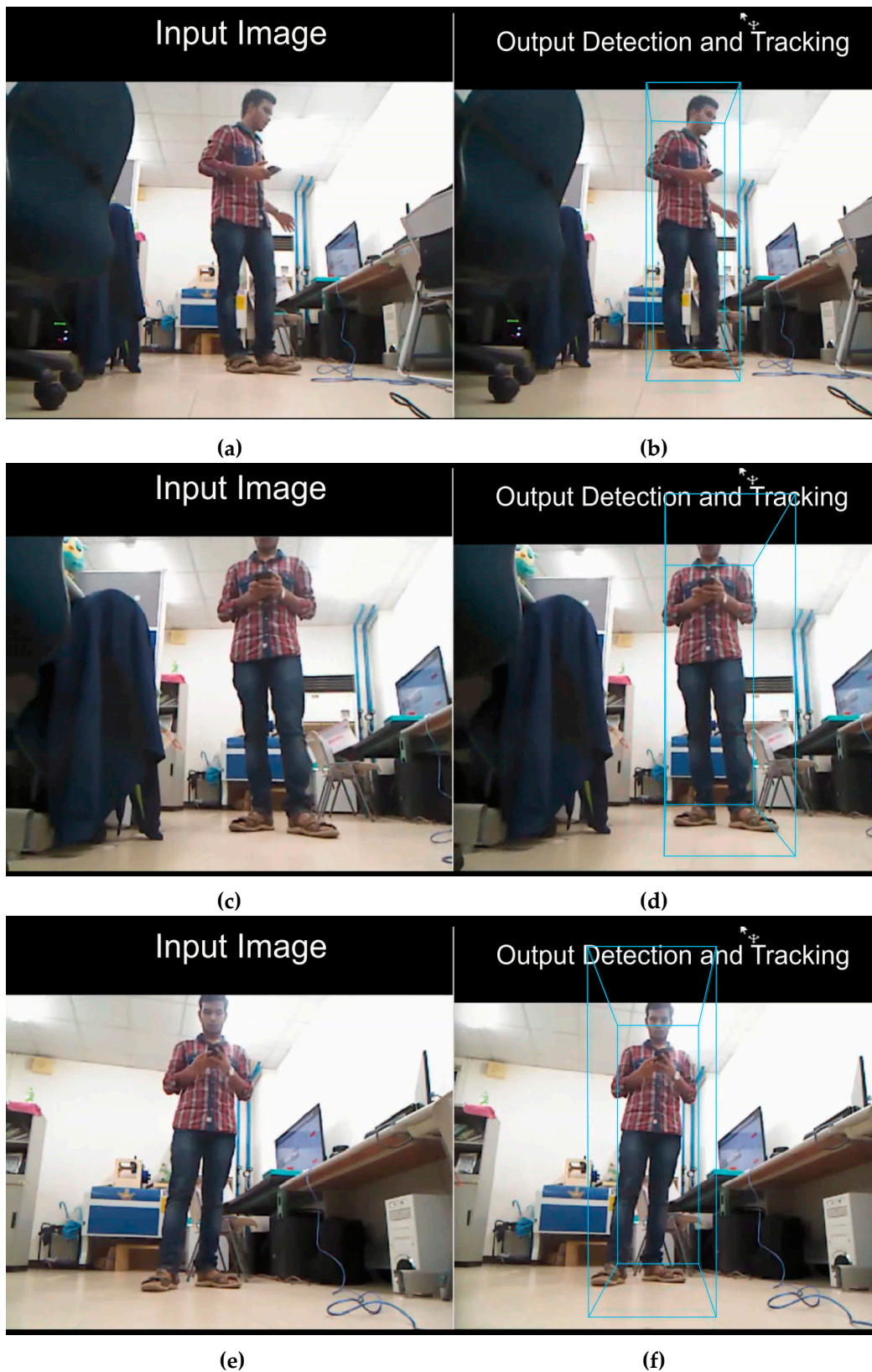
Figure 8. Recorded image sequence of interaction between one human and the mobile robot through wireless remote control.

The second runtime scenario is an experiment of detecting and tracking two humans in the room. Figure 9 shows the image sequence of detecting and tracking two humans using the proposed system

during the second experiment. Figure 9a,c is the input data captured from the Kinect, and Figure 9b,d show that two humans are detected and are tracked individually in two different colored polygons box in a single frame. Therefore, the above experimental results validate the performance of the proposed human detection and tracking system combined with the wireless remote-controlled mobile platform. A video clip of the above experimental results can refer to the website of reference [18].
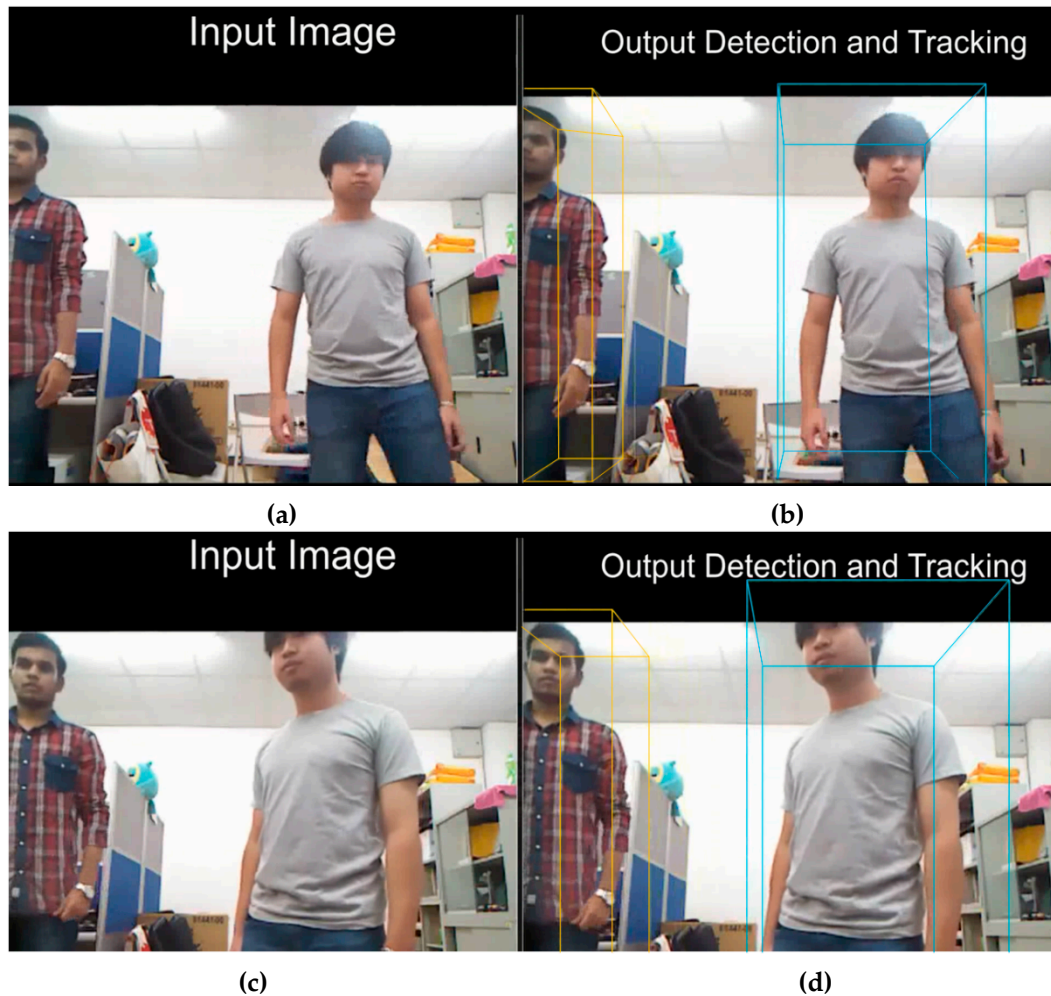

Figure 9. Recorded image sequence of detecting and tracking two humans in the room.

6. Conclusions and Future Work

In this paper, we have implemented real-time human detection and tracking function on a wireless controlled mobile platform, which can be easily adaptable as an intelligent surveillance system for security monitoring in venues like railway stations, airports, bus stations, and museums, etc. Experimental results show that the user can easily use a cell phone to remote control the mobile platform for the detection and tracking of multiple humans in a real-world environment. This design can improve the security level of many practical applications based on a mobile platform equipped with a Kinect device. Moreover, it is also helpful for applications of remote monitoring, data collection, and experimental purpose.

In future work, we will try to implement the human detection and tracking on an ARM-based embedded platform, e.g., Rasberry Pi 3 [19], Beaglebone [20]. Running on an embedded platform helps in decreasing the system complexity and wireless transmission for good communication. Moreover, we can add more RGB-D cameras on the proposed wireless controlled mobile platform for 360 degrees video capturing of the environment.

References

1. Quigley, M.; Gerkey, B.; Smart, W.D. *Programming Robots with ROS*, 1st ed.; 1005 Gravenstein Highway North: Sebastopol, CA, USA, 2015.
2. Zalewski, M. *The Tangled Web*; 38 Ringgold Street: San Francisco, CA, USA, 2011; Chapter 3.
3. Hussein, M.; Abd-Almageed, W.; Ran, Y.; Davis, L. Real-time human detection, tracking, and verification in uncontrolled camera motion environments. In Proceedings of the 4th IEEE International Conference on Computer Vision Systems, New York, NY, USA, 4–7 January 2006; p. 41.
4. Lee, J.; Choi, J.; Shin, D.; Shin, D. *Multi-User Human Tracking Agent for the Smart Home*; Pacific Rim International Workshop on Multi-Agents: Guilin, China, 2006; pp. 502–507.
5. Kulkarni, R.K.; Wagh, K.V. Human tracking system. In Proceedings of the International Conference on Electronics and Communication Systems, Coimbatore, India, 26–27 February 2014; pp. 1–6.
6. Suhr, J.K.; Jung, H.G.; Li, G.; Kim, J. Mixture of Gaussians-Based Background Subtraction for Bayer-Pattern Image Sequences. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 365–370. [CrossRef]
7. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2013**, *374*, 2065. [CrossRef] [PubMed]
8. Welch, G.; Bishop, G. *An Introduction to the Kalman Filter*; University of North Carolina: Chapel Hill, NC, USA, 2001.
9. Sales, F.F.; Portugal, D.; Rocha, R.P. Real-time people detection and mapping system for a mobile robot using a RGB-D sensor. In Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics, Vienna, Austria, 2–4 September 2014; pp. 467–474.
10. Jafari, O.H.; Mitzel, D.; Leibe, B. Real-time RGB-D based people detection and tracking for mobile robots and head-worn cameras. In Proceedings of the IEEE International Conference on Robotics and Automation, Hong Kong, China, 31 May–7 June 2014; pp. 5636–5643.
11. Nair, M.R.; Deepa, D.; Aneesh, R.P. Real time human tracking system for defense application. In Proceedings of the International Conference on Next Generation Intelligent Systems, Kottayam, India, 1–3 September 2016; pp. 1–6.
12. Priyandoko, G.; Wei, C.K.; Achmad, M.S.H. Human following on ROS framework for a mobile robot. *SINERGI* **2017**, *22*, 77–82. [CrossRef]
13. Nasif, S.; Khan, M.A.G. Wireless head gesture controlled wheel chair for disable persons. In Proceedings of the IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, Bangladesh, 21–23 December 2017; pp. 156–161.
14. Patil, M.; Abukhalil, T.; Patel, S.; Sobh, T. UB swarm: Hardware implementation of heterogeneous swarm robot with fault detection and power management. *Int. J. Comput.* **2015**, *14*, 1–14.
15. Zhang, Y. Support vector machine classification algorithm and its application. In Proceedings of the International Conference on Information Computing and Applications, Chengde, China, 14–16 September 2012; Volume 308, pp. 179–186.
16. Clarkson, K.L. *Nearest-Neighbor Searching and Metric Space Dimension*, 2nd ed.; MIT Press: Murray Hill, NJ, USA, 2005; Chapter 5.
17. Nevonsolutions Pvt Ltd. Four-Wheel Robotic Vehicle Chassis. Available online: https://nevonexpress.com/4-Wheel-Robotic-Vehicle-Chassis.php (accessed on 29 January 2019).
18. Experimental Results of the Proposed System. Available online: https://www.youtube.com/watch?v=AL9yPiNswDw (accessed on 29 January 2019).

19. Andrews, R. *Raspberry Pi for Beginners*; Richmond House: Richmond Hill, BC, Canada, 2014.
20. Molly, D. *Exploring Beagle Bone*; 10475 Crosspoint Boulevard: Indianapolis, IN, USA, 2015.