Embedded Human-Following Mobile-Robot with an RGB-D camera

Minh-Quoc Do, Chang-Hong Lin
Department of Electronic and Computer Engineering
National Taiwan University of Science and Technology
M10202805, chlin@mail.ntust.edu.tw

Abstract

Embedded device recently becomes more crucial in some real-life applications, such as the entertainment, health care, and industrial application. Given the development of embedded devices, the mobile robot becomes not only cheaper, but also more highly practical to produce. However, the embedded device still has some limitation from the hardware capability, therefore the implemented method has to be robust and low computational complexity. In this research, we introduce an embedded human following mobile robot application with an RGB-D camera, which can help us detect human position based on the human geometry, compatible into the ARM-based embedded device. The Human Depth of Interest (HDOI) is proposed to cooperate with CAM-shift algorithm for the human detection and tracking. The virtual spring model with the safe region and active region is chosen to carry out the smoothing navigation and easy implementation. The performance of our approach is enhanced more accuracy than the original CAM-shift algorithm and more feasible than the stereo vision system.

1. Introduction and Related Works

In recent years, it was experienced that the mobile robots have never failed to attract the engineer's attention. Some adopt that the human robot interaction is the integral part of the mobile robot research which has the high practical feature in human life. In the near future, with the communication with the robot, the human can give the commands to control robot, therefore the robots can be deemed as the equipment to change the human labor, through producing the human-friendly robot. One of the human robot applications, the human following function is one of the most popular services that we can easily see in the real life. Even though it is popular in human life, to find a robust method for human-following robots in real time is still a rough task.

Some researches install one single Laser Range Finder (LRF) [1] on the mobile robot to detect and tracking the human's legs through analyzing the reflect laser signal from the objects. Despite of the large detected range of the LRF, it still has the some disadvantages, such as high cost, limited distance measurement or the interferences. To avoid these problems, other researchers proposed an LRF with vision system, such as the human face or body detection systems as demonstrated in [2].

In other hand, the stereo vision [3, 4] is considered as the main sensor to detect and track human appearances. Based on the calculated distance from the two cameras, these systems can acquire the color and the depth information at the same time. Furthermore, Jia *et al.* [5] utilized the stereo camera with radio-frequency identification (RFID) to avoid the situation when the target leaves the camera view. However, the performance of stereo vision is highly influenced by the calibration parameter, the image processing computation, and especially the quality of the captured image in each camera.

One of the innovations of the stereo vision, the low cost RGB-D camera supports us not only the color image, but also the depth information. In this research, in order to tackle the problem in human detection and tracking with the advantage of embedded devices, we propose Human Depth of Interest (HDOI) to localize the human using the RGB-D camera, and improve the performance of CAM-shift tracking method. In addition, we modify virtual spring model with the safe and active region to reduce time consumption and make the proposed method more flexible in the embedded application.

The rest parts of this paper are organized as follows. Section 2 shows the system structure of this research. The vision method and following solution are explained in Section 3. Then, the experimental results are issued in Section 4, and the last one illustrates the conclusion.

2. System Overview

In this section, we present the hardware architecture of the mobile robot and the overview of the human following method of this paper.

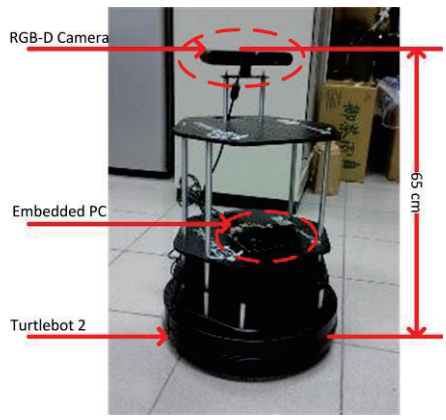

Figure 1. The mobile robot's hardware system

2.1. Mobile Robot Architecture

The Turtlebot-2 from Clear-path Robotics Inc. is used as the experimental platform, which has 7.6 cm diameter drive wheels and the weight is about 6.3 kg. The Robot Operation system (ROS) [6] is implemented in our research, due to its good support in ARM-based platform and propitious to manage sensors' data of the robot.

The UDOO Embedded PC with some features, such as CPU Quad-core 1.5GHz, 2GB of memory, one wireless module, and one integrated Arduino onboard, is applied as the main processing computer. Moreover, this PC is installed with the Ubuntu 12.04 OS and ROS Groovy ARM version. The ASUS Xtion Pro Live RGB-D camera is the main sensor of our research and is mounted at 65 cm from the ground as shown in Figure 1.

2.2. Overview of Human Tracking and Following

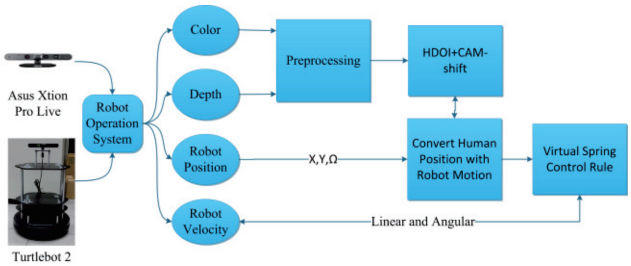

Figure 2. Flowchart of Human Following Method

Figure 2 depicts the flowchart of our mobile robot system. Firstly, The ROS is used in this application to support not only the color and depth information but also the velocity and position of the Turtlebot-2 under nodes. The depth information is utilized to generate the 3D world coordinates. These coordinates become the main role in human localization by using the body geometry detection. After estimating the human position, we apply the CAM-shift tracking algorithm adaptively in the next color frame with our proposed removing background method from this human position. Finally, the coordinate of human from the tracking step is used to calculate as the input parameters of virtual spring model, and the outputs of this control rule are the linear and angular velocity of robot.

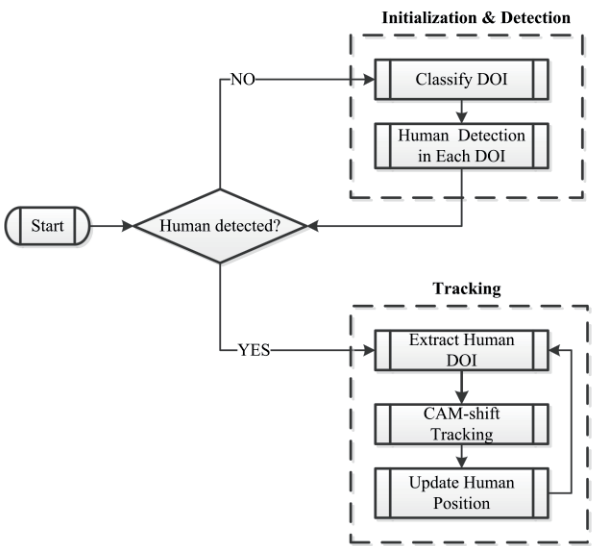

Figure 3. Human Detection and Tracking Model

3. Proposed Method

3.1. Preprocessing

In this step, due to the separation when acquiring the color and depth images of RGB-D sensors, therefore these cameras must be calibrated and established mapping accurately with the intrinsic parameters, such as focal length, distortion coefficients and image centers. Then, the resolutions of both color and depth cameras are set up as 320x240 to reduce the computation.

Then, depth information is processed by removing the noise, null pixels, calibrated before being used to calculate 3D coordinates. After that, the random sample consensus (RANSAC) algorithm is applied into Plan Surface Detector (PSD) method [7] to estimate the ground plan with initial distance from robot head to ground is 65 cm and remove the pixels belong to this plan in both color and depth images.

3.2. Human Detection and Tracking

The proposed human detection and tracking model is divided into two parts, as shown in Figure 3. The first part is for the initialization and human detection. In this part, the distribution of depth image is estimated through the histogram. Then, we follow the method presented by Zhang *et al.* [8] to classify the depth and color image into many Depth of Interests (DOIs) with the interval of 0.5 meters. With each DOI, the finding contour is implemented to separate the objects, three points (x_1, x_2, x_3) are randomly selected for each object under the 3D-coordinate to compute the norm vector $n[x, y, z] = (x_2 - x_1) \times (x_3 - x_1)$. The objects are recognized as human unless satisfy the below constrains in Table 1.

Table 1. Non-human Recognition Constrains

Constrain	Explanation
$y \approx 0$	Vertical plan, Ex: a wall…
$x \approx 0,\ z \approx 0$	Along y-axis, Ex: a table, a desk…
Width-based	Width $>$ Max human width
Height-based	Height $\notin$ [Min, Max] human height

The second part focuses into the human tracking with the robot motion. The original CAM-shift algorithm performs in color-based, therefore it is easy to get wrong results due to the similarity between the background and target. To solve this problem, we issue the HDOI method to cooperate with CAM-shift. The proposed HDOI is assumed as a DOI which contains the human. Grounded on the average human velocity and the processing time of CAM-shift, the detected human's position, after projecting in Section 3.3, is proposed as the depth center value to extract the new HDOI with the interval of 0.5 meters. The background of color image is ejected with the above extracted HDOI, therefore the CAM-shift could track human more accurately.

3.3. Human Position Projection with Robot Motion

In this research, we collect the robot position from the ROS odometry sensor node with the origin at the starting engine position of the robot. However, the 3D-coordinates

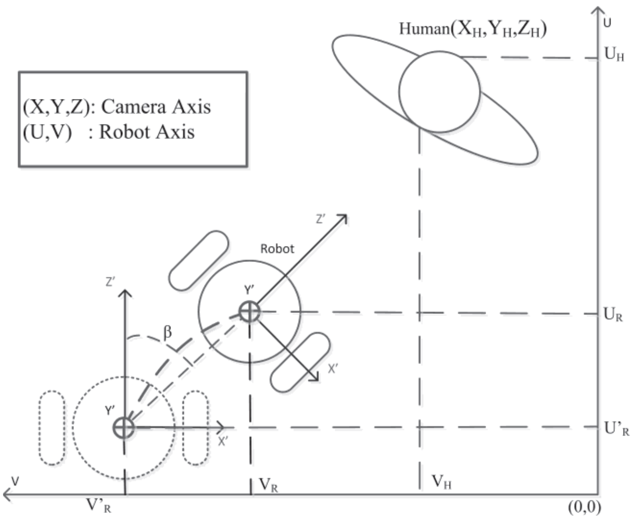

Figure 4. Human position under robot axis and camera axis

computed from RGB-D camera images have the origin at the placement of cameras on the mobile robot, therefore we require a coordinate projection related with the robot motion, as illustrated in Figure 4.

To avoid the mistake when extracting the HDOI, the Eq(1) is implemented to project the previous human position onto new robot coordinate.

$$\Delta V_R = V_R - V'_R$$

$$\Delta U_R = U_R - U'_R \quad (1)$$

$$X_{H\_new} = (-Z_H + \Delta U_R)sin\beta + (X_H + \Delta V_R)cos\beta$$

$$Y_{H\_new} = Y_H$$

$$Z_{H\_new} = (Z_H - \Delta U_R)cos\beta - (X_H + \Delta V_R)sin\beta$$

where ΔV_R and ΔU_R are the movement of robot, β is the robot's rotation angle compared with the previous position,($X_{H\_new}, Y_{H\_new}, Z_{H\_new}$) is the result of the projection of the previous human position (X_H, Y_H, Z_H) under the robot motion. This position after the projection is able to apply to extract the new HDOI.

3.4. Modified Virtual Spring Control Rule

From the detection and tracking method, we can get the human world coordinates in Eq(2).

$$W_H = \{x_H, y_H, z_H\} \quad (2)$$

The distance d from human to robot and the deviation angle θ between the virtual spring and robot are calculated from the above coordinates. Given the virtual spring model in [9], and the low computational requirement of embedded devices, Figure 5 illustrates the dynamic model in this research. The safe region is proposed as the arc area with the angle Φ bounded between d_1 and d_2, which are determined by the experiments, and the remaining region is called the active region. We assume when he or she stays in the safe region, robot is difficult to lose the human target, therefore the control rule is not applied in this situation. Otherwise, in the active region, the control rule is computed by Eq(3) and Eq(4).

$$F_1 = k_1(d - d_o) \quad (3)$$
$$F_2 = k_2\theta$$

where k_1, k_2 are the expansion (N/m) and bending (N/rad) coefficients, d_0 is the distance at the equilibrium state, F_1 is the expansion force and F_2 is the bending force from the virtual spring model.

$$m\dot{v} = F_1 \cos\theta + F_2 \sin\theta - k_3 v \quad (4)$$
$$I\dot{\omega} = (F_1 \sin\theta + F_2 \cos\theta - k_4\omega) * L$$

where v, ω is linear velocity (m/s), and angular velocity (rad/s), k_3,k_4 are the viscous fiction coefficients (N.s/m and N.s/rad), m is the total mass of the robot(kg), I is the inertia moment, and L is the distance from center of robot to the joint (m).

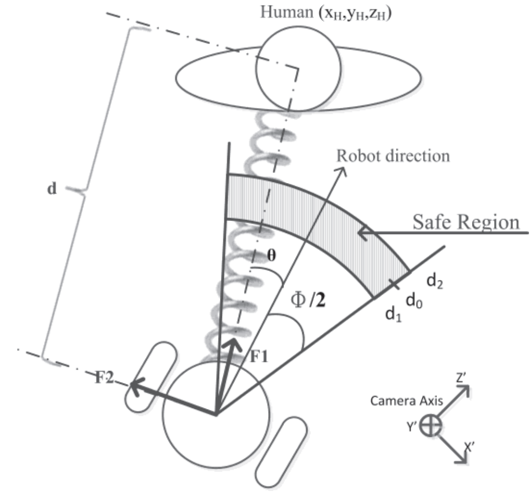

Figure 5. Proposed Virtual Spring Model

4. Experimental results

Table 2. Parameters Setup

Parameters	Value
Equilibrium distance d_0 (m)	1.95
Expansion coefficient k_1 (N/m)	1
Bending coefficient k_2 (N/rad)	4
Fiction coefficient on linear k_3 (N.s/m)	3
Fiction coefficient on angular k_4 (N.s/rad)	3
Moment of Inertia I	0.394
Distance from center of robot to the joint L (m)	0.354
Upper distance safe region d_2 (m)	2
Lower distance safe region d_1 (m)	1.7
Angle of safe region arc Φ (rad)	0.1

In the experiment, the C++ with the OpenCV library [10] is used to implement the proposed method. The control parameters follow as Table 2. Given the impacts of background, the original CAM-shift misses the human target, while the proposed method still tracks correctly, as be illustrated in Figure 6.

Table 3. Latency Comparison

Method	Latency(ms)
HOG [11]	224 to 230
Stereo vision + EKF [3]	~100
HDOI+ CAM-shift	26.8 to 39.1

In Table 3, the latency of our approach and the Histogram of gradient (HOG) [11] are measured on a PC with Intel Core i3 @ 2.13 GHz, 4GB of RAM, and Stereo

Vision with Extend Kalman Filter (EKF) of Petrović *et al.* [3] is selected to compare due to the similarity in using the embedded PC. In their research, they used a low power PC to acquire images, and processed images of cameras in a PC with Intel Xeon E5520 @ 2.26 GHz, 6GB of RAM.

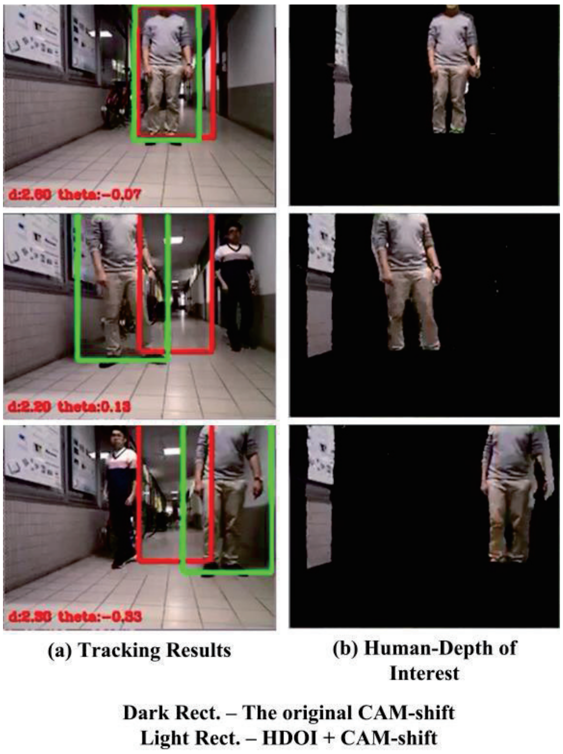

Figure 6. The original CAM-shift and HDOI+ CAM-shift

The following behavior of the mobile robot is captured under the other camera view in an indoor environment. The proposed situation of this experience is that the human walks, stop, go backward and then human continues to turn left at T-corner of corridor. Follow as Figure 7, the first detected human (or the nearest human) is chosen as the person to be followed at (a), robot begins to follow (b), the robot stops when it recognizes human is staying and stops in the safe region (c), then robot moves backward when human approaches it (d), continues going straight (e), and turns left to follow human at a T-corner (f).

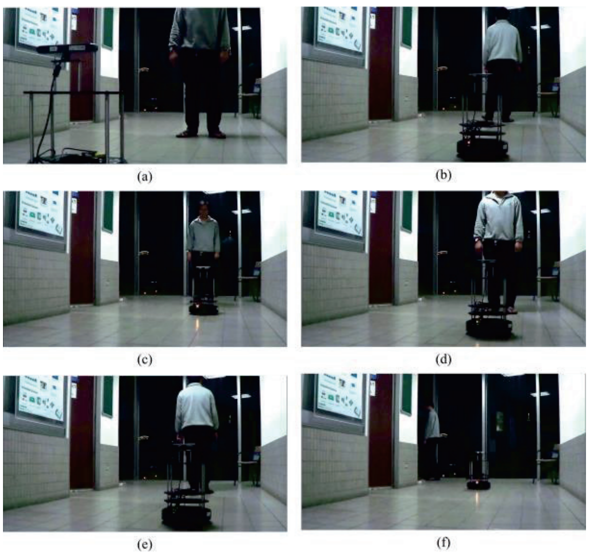

Figure 7. Human Following Results

5. Conclusions

In this paper, we proposed a low computation method for the embedded ARM-based following mobile robot. The DOI is improved into the HDOI to be suitable to develop in the human following mobile robot application. Moreover, the background is removed by using the HDOI algorithm, and result into eliminating the impact of the environment on the CAM-shift approach. The virtual spring model with the safe and active region makes the robot more safe and workable when following human. In addition, the experimental results show us the processing time of the proposed method is better than some prior arts, such as the HOG [11] and the Stereo Vision with EKF [3]. In the future works, the combination with other sensors is considered to make the assistive robot much friendlier.

Acknowledgement

This work was supported by the Ministry of Science and Technology of Taiwan, R.O.C, under project 103-2220-E-011-002-.

References

[1] W. Chung, H. Kim, Y. Yoo, C.-b. Moon, and J. Park, "The Detection and Following of Human Legs Through Inductive Approaches for a Mobile Robot With a Single Laser Range Finder," *IEEE Trans. Ind. Electron.,* vol. 59, pp. 3156-3166, 2012.

[2] N. Bellotto and H. Hu, "Multisensor-Based Human Detection and Tracking for Mobile Service Robots," *IEEE Trans. Syst., Man, Cybern, B, Cybern. ,* vol. 39, pp. 167-181, 2009.

[3] E. Petrović, A. Leu, D. Ristić-Durrant, and V. Nikolić, "Stereo Vision-Based Human Tracking for Robotic Follower," *Int. J. Adv. Robot Syst.,* vol. 10:230, 2013.

[4] J. Satake and J. Miura, "Multiple-Person Tracking for a Mobile Robot using Stereo," in *Proc. IAPR Conf. on Mach. Vision App.*, 2009, pp. 273-277.

[5] S. Jia, L. Wang, S. Wang, and C. Bai, "Fuzzy-based intelligent control strategy for a person following robot," in *Proc. IEEE Int. Conf. ROBIO*, 2013, pp. 2408-2413.

[6] ROS-Introduction: http://wiki.ros.org/ROS/Introduction

[7] R. Cupec, R. Grbić, E. K. Nyarko, K. Sabo, and R. Scitkovski, "Detection of Planar Surfaces Based on RANSAC and LAD Plane Fitting," in *Proc. 4th European Conf. on Mobile Robot*, 2009.

[8] H. Zhang, C. Reardon, and L. E. Parker, "Real-Time Multiple Human Perception With Color-Depth Cameras on a Mobile Robot," *IEEE Trans. Cybern.,* vol. 43, pp. 1429-1441, 2013.

[9] K. Morioka, J.-H. Lee, and H. Hashimoto, "Human-following mobile robot in a distributed intelligent sensor network," *IEEE Trans. Ind. Electron.,* vol. 51, pp. 229-237, 2004.

[10]D. G. R. Bradski and A. Kaehler, *Learning opencv, 1st edition*: O'Reilly Media, Inc., 2008.

[11]N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2005, pp. 886-893 vol. 1.