A Real-time Moving Target Following Mobile Robot System with Depth Camera

Xueling Luo¹, Dan Zhang¹,* and Xiaodong Jin¹,²

¹ York University, Toronto, Canada, M3J 1P3
² Beijing Jiaotong University, Beijing, China, 100044

*dzhang99@yorku.ca

Abstract. An intelligent system that could obtain a certain moving target human and could follow the target in real-time by a mobile robot could be widely applied in manufacturing as assistant or service industry to provide a human robot interaction experience. In this work, an improved efficient human following algorithm for a mobile robotic system is proposed. The system consists by a mobile robot and a depth camera which overcomes target scale variation, missing and occlusion. An improved three-dimensional tracking method is presented with the depth camera. Besides, the depth camera also provides information to avoid obstacles in robot’s path. The omnidirectional mobile robot in the system provides a flexible and fast motion responding solution. In the experiments, the proposed system is tested in different scenes and the results show the reliability of the 3D tracking system in terms of accuracy and robustness to the environment.

1. Introduction

Human tracking is one of the most important and promising application for a robotic system in the service industry and manufacturing. A human tracking robot could play an important role in daily life such as a cargo transportation assisting robot at the workshop or an individual luggage carrying robot at any transportation terminal [1]. Therefore, a human following robotic system should able to detect the target continuously and keep follow movement of the target while could acknowledge surrounding environment to find its own path. Unlike surveillance systems, cameras are not static in human tracking robotic system. Challenges for specified moving target localization in a mobile robotic system are brought by environment changing during the movement such as camera position changing, illumination variation, changing of the target shape and target occlusion. [2]

In recent years, researches have put effort on moving target detection and tracking for human robot interaction application [3]. Several researches have been presented to achieve human following for a robotic system with different sensor sets approach such as using laser scanner to discriminate human legs with other objects [4] and moving targets tracking with integrated range sensors [5]. Also, several studies present on path planning and localization to achieve moving target following [6]. Applying visual system for moving target tracking is one the most popular research area among them. Texture features usually been applied on targets localization in a series of RGB images and those features could be generally classified into two categories: global features and local features. To consider the efficiency requirement on real-time tracking system, a few remarkable work have been present by using local feature in visual tracking filed [7]-[9]. In reality world, certain moving target tracking is a three dimensional problem and it is different from general pedestrians tracking problem that all human
in the picture are targets. Discrete Fourier Transform (DFT) is one of the typical methods have been used on solving correlation filter to achieve high calculating efficiency [10]. However, such methods are using for two dimensional image processing.

In this paper, we present an efficient and accurate moving target detecting and tracking algorithm for a mobile robot which equip with an RGB-D Camera and a laptop. An efficient modified Kernelized Correlation Filter (KCF) is applied as visual tracker. Information obtained from depth camera is used to calculate 3D position of the target and provides environmental information for the robot to decide its path. The system has great performance on the target detecting, tracking and following and the system is able to moving with avoiding obstacles if there are some in the scene.

2. Proposed method

In this section, we will briefly introduce the KCF tracker which is the base of our tracking algorithm. A lot of KCF trackers are designed for searching target in a window without drift and tracking without failure when the target is moving with changing speed. To achieve that more calculation is needed which highly reduce computing speed of the algorithm which is not applicable in real-time tracking system. [11] We take variation velocity of the target in to consideration to modify the KCF tracker and to use the same vector of the movement for the mobile robot to predicting target trajectory. Besides, we also include the depth camera information in the algorithm for the mobile robot to avoid obstacles in its path.

2.1. Kernelized correlation filter

The goal of Kernelized Correlation filter (KCF) tracker [7] is to solve the ridge regression which is to minimize error between input image patches $t_{w,h}$ ($w \times h$) and its regression labels $y_{w,h}$.

$$\min \sum_{w,h}|f(t_{w,h}),W) - y_{w,h}|^2 + \gamma ||W||^2$$  \hspace{1cm} (1)

Where $f$ represents the mapping to the nonlinear space which introduced by kernel, $t_{w,h}$ is the image patch of the target, the $\gamma$ is regular parameter to control overfitting, $W$ So, the target has to be selected manually at beginning. The shifts $t_{w,h}$ ($(w,h) \in \{0, ..., W-1\} \times \{0, ..., H-1\}$) around the target are considered as training sample. $W$ is the solution of $W = \sum_{w,h} \alpha_{w,h} f(t_{w,h})$.

The variable is presented as calculating discrete Fourier transformation and its inverse of following expression:

$$\alpha = F^{-1} \left[ \frac{f(y)}{F(K(t_{w,h},t)) + \gamma} \right]$$  \hspace{1cm} (2)

Where $K(t_{w,h},t)$ represents the kernel correlation of the target model $t$. Then the target position within a pitch $z$ could be found by the maximum response of

$$\hat{f}_z = F^{-1} \left( F \left( K(t_{w,h},t) \right) F(\alpha) \right)$$  \hspace{1cm} (3)

2.2. Target tracking

In order to tracking a target without constant moving speed and to predict the motion, the system variable vector is defined as: $x = [p_x, p_y, v_x, v_y]^T$. Where $(p_x,p_y)$ is the centre location of the target and $(v_x,v_y)$ is the velocity of the target. The initial position of the target need to be selected manually and initial velocities are zero. To use a-second order autoregressive model for target status estimating [11]:

$$x_{k+1} = Ax_k + N(\tau)$$  \hspace{1cm} (4)

Where A is the state matrix and $N(\tau)$ is the noise matrix which is Gaussian distributed.

To improve tracking accuracy, the system is described as follow for three dimensional position prediction as $x_p = [p_x, p_y, p_z, v_x, v_y, v_z]^T$, where $(p_y)^2 = (d_y)^2 - (p_x)^2 + (p_z)^2$ is the depth position of the target respect to the mobile robot which is obtained by the depth camera and $d_x$ is the
depth of position x. Position offset from x direction to the z direction could be represent by 
\[ \sin \theta_r r = d_y/p_x \] where \( \theta_r \) is the rotation vector for mobile robot control. And \( v_y = d (p_y) \) is the moving velocity of the target in depth direction. We applied extended Kalman filter for target position predicting and following. The target state and measurement equations could be written as

\[ \hat{X}_{k+1} = \hat{X}_k \]  
\[ P_{k+1} = \phi_{k+1}P_k \phi_{k+1}^T + Q \]  

Where \( P_{k+1} \) is the covariance matrix and \( \phi_{k+1} \) is the system. Q is the noise of the system. The state of system would be updated by

\[ K_{k+1} = P_{k+1}H_{k+1}^T (H_{k+1}P_{k+1}H_{k+1}^T + R)^{-1} \]  
\[ \hat{X}_{k+1} = \hat{X}_{k+1} + K_{k+1}(Z_{k+1} - \hat{h}(X_{k+1})) \]  
\[ P_{k+1} = [I - K_{k+1}H_{k+1}]P_{k+1} \]  

Where \( \hat{h}(X_{k+1}) \) is the predicted measurement model of the system and \( Z_{k+1} \) is the measurement result from the system.

2.3. Obstacle avoiding

In this work, only relative position of obstacles in the camera view are need to be determinated. The depth camera we used in this work is Kinect which could provide RGB and depth images synchronously. When there is no obstacle in the scene, the mobile robot moves in vector direction \([X_r, Y_r, \theta_r, V_r]^T\) with \([X_r, Y_r, \theta_r]\) as moving direction which are obtained from prediction in 2.2 and \( V_r \) as moving velocity which is calculated from \( v_x \) and \( v_y \) since the robot is not able to move in vertical direction. In figure 1, the module of tracking system is shown when \( p_x = 0 \).

![Figure 1. The system coordinate](image1)

When there are obstacles in the path of the robot. In order to receive abnormal notice timely during the mobile robot moving, we only consider information from the depth camera in the maximum height of the robot \( H_{r, max} \). In figure 2, a detection result is presented with depth camera in robot height. All marked point in the picture shows the edges of obstacles in the scene for accessibility judgement. We assume the robot could pass through an alley with minimum width \( W_{min} \) or it will stop. So the robot will have an adjusting vector \([x_r, y_r, \rho_r, v_r]^T\). Where \( x_r = -x_o \) and \( x_o \) is the location in x direction of the obstacle, \( y_r \) will be the same as \( Y_r \), \( \sin \rho_r = f_1(X_r - x_r, Y_r) \) and \( v_r = f_2(X_r - x_r, Y_r) \).

![Figure 2. The detection result from depth camera with robot height \( H_{r, max} \)](image2)

2.4. Mobile robot

Omnidirectional robot is able to move in more direction comparing with normal wheeled robot. Different motions of the system are decoupled for a simple motion like translation or rotation which
provide flexibility and fast responsive performance for the system. [12] The mecanum wheeled robot is a typical omnidirectional system which is shown in Figure 3.

**Figure 3.** The configuration of the robot

Where, the coordinates OXY is the fixed coordinate and the $O_rX_rY_r$ is the moving coordinates for the robot. Besides, the coordinates $o_{wi}$ ($i=1, 2, 3, 4$) are coordinate frame for each wheel.

Let the position of each wheel to be described as $S_{wi} = [x_{wi}, y_{wi}, \alpha_i]^T$ ($i=1, 2, 3, 4$), $\hat{\theta}_{xi}$ is the angular velocity around the hub, $\hat{\theta}_{ri}$ is the angular velocity of the roller and $\hat{\theta}_{zi}$ is the angular velocity about the contact point. $R$ is the radius of the wheel and $r$ is the roller radius. $\varphi_i$ is the slope angle for the roller and in the robot, $\varphi_1 = \varphi_3 = 45^\circ$ and $\varphi_2 = \varphi_4 = -45^\circ$. $\dot{S}_{wi}$ could be represented as equation (10).

$$\dot{S}_{wi} = \begin{bmatrix} \dot{x}_{wi} \\ \dot{y}_{wi} \\ \dot{\alpha}_i \end{bmatrix} = \begin{bmatrix} 0 \\ R \end{bmatrix} \begin{bmatrix} r \sin \varphi_i \\ -r \cos \varphi_i \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} \dot{\theta}_{xi} \\ \dot{\theta}_{ri} \end{bmatrix} = \begin{bmatrix} 0 \\ r \sin \varphi_i \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} \dot{\theta}_{xi} \\ \dot{\theta}_{ri} \end{bmatrix}$$

(10)

The coordinates translation matrix from wheel coordinates to robot coordinates $T_{wi}$ is given as

$$T_{w2ri} = \begin{bmatrix} \cos \alpha_{w2ri} & -\sin \alpha_{w2ri} & d_{xw2ri} \\ \sin \alpha_{w2ri} & \cos \alpha_{w2ri} & d_{yw2ri} \\ 0 & 0 & 1 \end{bmatrix}$$

(11)

Where $\alpha_{w2ri}$ ($i=1,2,3,4$) is the rotation angle of wheel to the $O_rX_rY_r$ coordinates, $d_{xw2ri}$ and $d_{yw2ri}$ are the translational distance between two coordinates. For the position of the robot, $S_r = [x_r, y_r, \alpha_r]^T = T_{w2ri}S_{wi}$. In figure 2, $\alpha_{w2ri}$ could be found equal to 0.

Then, the inverse kinematics is obtained as

$$\begin{bmatrix} \dot{x}_r \\ \dot{y}_r \\ \dot{\alpha}_r \end{bmatrix} = \frac{R}{4} \begin{bmatrix} -1 \\ 1 \\ 1 \\ W+L \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \\ \dot{\theta}_4 \end{bmatrix}$$

(12)

Where $[\dot{\theta}_1, \dot{\theta}_2, \dot{\theta}_3, \dot{\theta}_4]^T$ is the angular velocity of each wheel.

In the control of the robot, the signal could be sent in as the coordinate $[\dot{x}_r, \dot{y}_r, \dot{\alpha}_r]$.

3. Experiment

In figure 4, detection and tracking results are shown. During the tracking, the algorithm is able to track the target when the target is occluded and lost for a short time. When the target is partially occluded and fully occluded, the target box is still located on the target path. With relatively long tracking period with complex interference, the target could still be tracked.
In the mobile robot system used in the experiment is shown in figure 5 which consists by an omnidirectional mobile robot, a Kinect camera system and a laptop. The camera system is already made into wireless power supply. The experiment is run in MATLAB 2018a.

In figure 6, few pictures from the experiment is shown.

4. Conclusion
In the work, we use a relatively simple system to achieve certain human tracking algorithm for a mobile robot in 3D. To improve efficiency and accuracy of the system, a modified KCF is applied as the target tracker and a EKF is also used to update information from different dimensional. The system is verified in a real system and achieves functions as detecting, tracking the target and control the mobile robot following the target. However, when the target is moving too fast with changing speed the system is not perfectly worked for the mobile robot following part. We will keep put our effort on it in our future work.

Acknowledgement
The authors would like to thank the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors gratefully acknowledge the financial support from Kaneff Research Chairs program.
References

[1] Hirose N, Tajima R, and Sukigara K, “Personal robot assisting transportation to support active human life—Human-following method based on model predictive control for adjacency without collision,” in Proc. IEEE Int. Conf. Mechatronics (ICM), Nagoya, Japan, 2015, pp. 76–81.

[2] Gupta M, Kumar S, Behera L, and Subramanian V. K., “A Novel Vision-Based Tracking Algorithm for a Human-Following Mobile Robot,” IEEE Trans. on Systems, Man, and Cybernetics: Systems, vol. 47, no. 7, 2017, pp. 1415-1427.

[3] Bellotto N and Hu H, “Multisensor-based human detection and tracking for mobile service robots,” IEEE Trans. Systems, Man, and Cybernetics, B, Cybernetics., vol. 39, no. 1, 2009, pp. 167–181.

[4] Xiao Q, Sun F, Ge R and et al., “Human Tracking and Following of Mobile Robot with a Laser Scanner,” 2017 2nd Int. Conf. on Advanced Robotics and Mechatronics (ICARM), Anhui, China. 2017, pp. 675-680

[5] Hoshino F and Moriya K, “Human Following Robot based on Control of Particle Distribution with Integrated Range Sensors,” IEEE/SICE Int. Symposium on System Integration, Kyoto, Japan, 2011, pp. 212-217

[6] Yoshimi T, Nishiyama M, Sonoura T, et al., “Development of a Person Following Robot with Vision Based Target Detection,” IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 2006, pp. 5286-5291.

[7] Henriques J. F., Caseiro R, Martins P, and et al., “High-speed tracking with kernelized correlation filters,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 37, no. 3, 2015, pp. 583–596.

[8] S. Liu, T. Zhang, X. Cao, and C. Xu, “Structural correlation filter for robust visual tracking,” in Proc. IEEE Conf. Computer Vision Pattern Recognition (CVPR), Las vegas, United States, 2016, pp. 4312–4320.

[9] T. Liu, G. Wang, and Q. Yang, “Real-time part-based visual tracking via adaptive correlation filters,” in Proc. IEEE Conf. Computer Vision Pattern Recognition (CVPR), Boston, United States, 2015, pp. 4902–4912.

[10] Wang M, Liu Y, Su D and et al., “Accurate and Real-Time 3-D Tracking for the Following Robots by Fusing Vision and Ultrasonar Information,” IEEE/ASME Trans. On Mechatronics, vol. 23, no. 3, 2018, pp. 997-1006

[11] Wang X, O'Brien M, Xiang C and et al., “Real-time visual tracking via robust Kernelized Correlation Filter,” IEEE Int. Conf. on Robotics and Automation (ICRA), Singapore, 2017, pp. 4443-4448

[12] Taheri H, Qiao B and Ghaeminezhad N, “Kinematic Model of a Four Mecanum Wheeled Mobile,” International Journal of Computer Applications, vol. 113, no.3, 2015, pp. 6-9

[13] Ess a, Leibe B and Gool L. V., “Depth and Appearance for Mobile Scene Analysis,” IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007