2C-01

Simultaneous Recognition and Tracking of Multiple Users for Autonomous Guide Robot

（自律案内ロボットにおけるユーザグループの認識と追跡を同時に行うシステム）

Bin ZHANG† Tomoaki NAKAMURA† Masahide KANEKO†

The University of Electro-Communications†

1. Introduction

In recent years, the development of service robots has played an important role in human-centered robotics. One important function of service robots used in different fields is to guide users from one place to another. In order to know the status of the users for providing them better services, a method to recognize and track them simultaneously is proposed. The tracking problem is modeled as finding the maximizing a posterior (MAP) solution of a series of user status, and is solved by using Markov Chain Monte Carlo (MCMC) particle filter [1]. Each of the users is recognized by integrating his/her personal information of color and face, and the recognition results are used for calculating the observation likelihood for the tracking process. By this method, every user can be recognized and robustly tracked during the guide task.

2. Methods

The overview of our system is shown in Fig. 1.

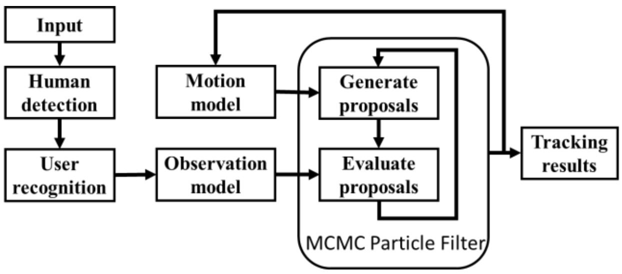


Fig. 1 Overview of the proposed system.

2.1 Human detection

The positions of the humans are detected by the backward Kinect sensor on the guide robot. The depth information with the human height range (1.0m-2.0m) is projected on the ground, and the human candidate areas are detected out after deleting the background parts. Connected-component labeling [2] process is used to find out the areas within human sizes by Eq. (1).

$$l_{min} \le l_{width} \le l_{max} \; ; \; l_{min} \le l_{length} \le l_{max} \qquad (1)$$

Simultaneous Recognition and Tracking of Multiple Users for Autonomous Guide Robot
†Bin ZHANG (UEC), Tomoaki NAKAMURA (UEC), Masahide KANEKO (UEC)

Here, l_{width}, l_{length} mean the width and length of a human candidate area, l_{min}, l_{max} mean the thresholds for a human area. The ranges of l_{width} and l_{length} are the same as the humans may face to different directions.

2.2 Recognition of particular users

The users are identified by integrating their personal features of face and color. The face feature can express a person well and the identification accuracy is very high by using the OkaoVision [3] software. The software can detect and identify the faces in an image, and meanwhile calculate the identification confidence. However, we cannot make sure that the face information is always available. The users may face to other directions or be occluded by others during the guide process, as they would move freely during the guide task. In this case, color information will be used for recognizing the person until his/her face is detected and identified again. The face and color information for each person is recorded during the initialization process. Here 5 frontal face photos are registered for each person, and we use color histogram of Hue and Saturation channels in HSV color space as the color information.

2.3 MCMC particle filter based tracking

The tracking problem can be formulated as finding the maximum posterior, and it is solved by using MCMC particle filter. We set the configuration of multiple users as X, which represents the locations, speeds and accelerations of the users, and the observations as Z, which represents the users' recognition results. Our aim is to find the maximum posterior of $p(X_t|Z_{1:t})$ following the Equation (2).

$$p(X_t|Z_{1:t}) \propto p(Z_t|X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|Z_{1:t-1}) dX_{t-1} \quad (2)$$

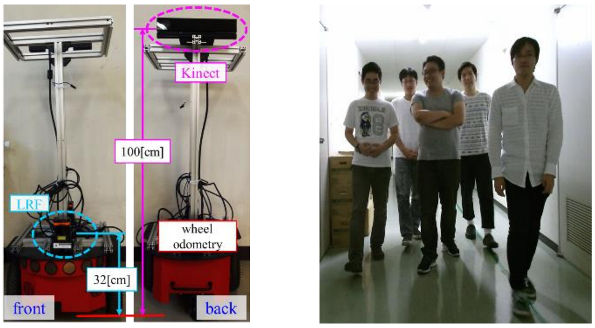
Here, $p(Z_t|X_t)$ means observation likelihood, which is calculated by the particular users' recognition results; $p(X_t|X_{t-1})$ means motion prediction, which is designed by Equation (3); and $p(X_{t-1}|Z_{1:t-1})$ means the prior, which is the posterior for the previous time step.

$$X_t = X_{t-1} + X'_t dt \; ; \; X'_t = X'_{t-1} + X''_t dt \; ; \; X''_t = X''_{t-1} + \mu \quad (3)$$

Here, μ is a process noise, and it is set as Gaussian noise.

3. Experiments

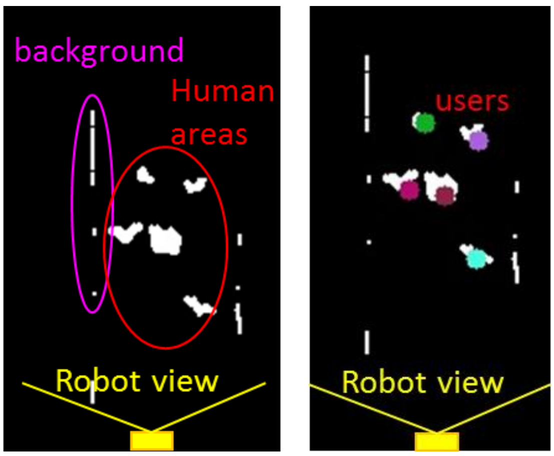
The guide robot that we used in our experiment is shown in Fig.2 (a). POINEER 3-DX is used as our mobile platform. One LRF sensor (UTM-30LX) is set in front of the robot (height: 32cm). A backward Kinect is set on the robot on the height of 100cm. One of the guide scene is shown in Fig.2 (b). The guide robot moves in front of the users to the destination.



(a) View of guide robot (b) Guide scene

Fig.2 The guide robot.

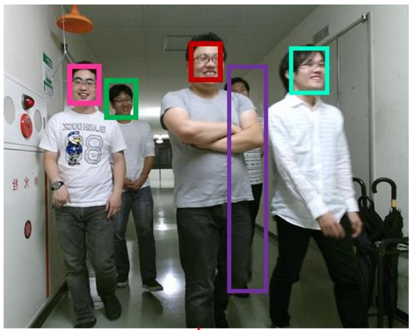
In the experiment, the robot guides a group with five users to get through a passage from a lab. For the scene shown in Fig.2 (b), the five users are detected by Kinect sensor, and the projection result of 3D distance information is shown in Fig.3 (a). The human areas (circled by red line) can be found out after background subtraction (circled by pink line) and area size threshold processing. The users are recognized by the face information. When the face cannot be detected, color information will be used. In the view shown in Fig.2 (b), the five users are successfully detected and recognized. The recognition results are shown in Fig. 3 (b). Five users' recognition results are shown as the colored circles. However, the users still may not be recognized correctly when they are occluded with each other or different people wear clothes with same color.



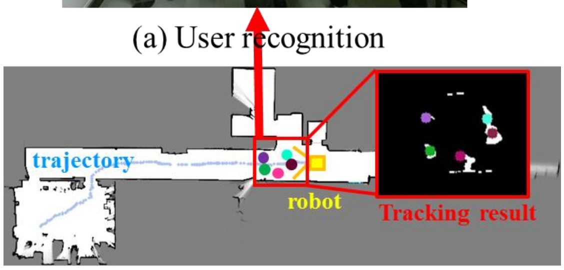
(a) Human detection (b) User recognition

Fig.3 The human candidate detection and user recognition result.

In this case, the tracking process will work to keep the trajectory smoothly. By using particle filter, the tracking results cannot jump far away so that some frames of improper recognition do not influence the result greatly. Moreover, even if the system fails to track the right person, it can quickly modify the tracking errors when the frontal face information of the user is detected and recognized in a high accuracy. The tracking result of the task is shown in Fig.4. The motion trajectories of the users on the map can also be gotten.



(a) User recognition



(b) User tracking result

Fig.4 The tracking results of users' group.

4. Conclusion

This paper proposed a simultaneous recognition and tracking method of multiple users group for an autonomous guide robot, and it is proved to work well by tracking multiple users groups under lab environment. Future work will be focused on improving the accuracy of particular users' detection and recognition rates and evaluating the effectiveness of the proposed method under public environments, like museums or airports.

References

[1] W. Choi, C. Pantofaru, and S. Savarese : "A general framework for tracking multiple people from a moving camera," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp.1577-1591, 2013.

[2] L. He, Y. Chao and K. Suzuki : "A run-based two-scan labeling algorithm," IEEE Trans. on Image Processing, vol. 17, no. 5, pp.749-756, 2008.

[3] E. Takigawa, S. Hosoi, M. Kawade : "Gender, Age, Human race identification & Estimation technology of facial attributes-OKAO Vision," Image Laboratory, vol. 15, no. 4, pp 82-87, 2004.