Autonomous Docking in a Human-Robot Collaborative Environment of Automated Guided Vehicles

Kai-Tai Song, Chien-Wei Chiu, Li-Ren Kang, Yu-Xuan Sun and Ching-Hao Meng
Institute of Electrical and Control Engineering, College of Electrical and Computer Engineering
National Chiao Tung University, Hsinchu, Taiwan R.O.C.
ktsong@mail.nctu.edu.tw, vi543210@gmail.com, arieskang1996.eed03@g2.nctu.edu.tw,
jk2626173jk.eed03@g2.nctu.edu.tw, johnson.meng73@gmail.com

Abstract— **In this paper we propose an autonomous docking and human-robot collaboration system for an automated guided vehicle (AGV). The AGV can not only navigate and dock autonomously, but also collaborate with the human by recognizing human in the environment. A human motion detection system is developed for the proposed human-robot collaboration design. A deep learning network is adopted to detect and recognize humans in the environment. By knowing of human motion, the AGV adjusts the automatic docking behavior in a collaborative manner. Practical experimental results demonstrate that human workers can co-exist with an AGV in an unstructured environment for autonomous docking tasks.**

Keywords—Autonomous Docking, Navigation System, Human recognition, Human-Robot Collaboration

I. Introduction

It has been noticed that high percentage of time for a work piece in a factory is actually spent in waiting or transferring among processing stations before it can be fabricated. Automated guided vehicles(AGVs) have become more and more popular in modern factories for material transferring. It plays an important role for an efficient production. On the other hand, collaborative robots(Cobots) are emerging in intelligent manufacturing in recent years. In a cobot working scenario, human workers are allowed to work together with robots in the same workspace. This increases the flexibility of manufacturing. Many useful tools have been developed for human-robot collaboration. But they are mainly for robotic arms, which are fixed in a production line. It demands urgent attention to develop a control framework for an AGV or mobile manipulator to work in a human-robot collaborative environment.

Therefore, if an AGV can detect and recognize humans in the environment, it will be able to work collaboratively with human workers. For face recognition, Taigman *et al.* [1] proposed the DeepFace architecture and Schroff *et al*. [2] developed the FaceNet. These are powerful tools to capture images of human faces and identify different people in the image. However, in the context of AGV applications, human faces are not visible at all times. Sometimes, the robot would need to recognize a specific person through body shape and personal characteristics. The Person Re-identification proposed by Almazan *et al.* [3] provides a useful tool for human recognition based on body features. Through the Re-id deep neural network, the detected person is compared with the person's image in the query to confirm whether the person is the same person. After identifying the specific person, the robot will be able to collaborate with him/her by knowing the human location.

Further, human motion information is essential for an AGV to interactive with a person. In this paper, we propose a method to acquire the important human-motion information for human-robot collaboration. Thus, the AGV can make smart decisions to interact with people during a docking process. It can dock near to a specific human for loading and unloading of materials. In the proposed method, in addition to enable the AGV to accurately dock at the station, the AGV can also interact with people during the docking process, and can stop near a specific human to provide service, while avoid other personnel in the task. The human-motion information can allow the AGV to schedule more flexibly, and can also complete a task more efficiently.

II. Autonomous AGV Docking System in an Human-Robot Collaborative Environment

The proposed system architecture for the Intelligent AGV is shown in Fig. 1. The proposed system consists of four parts: AGV localization, path planning, docking control and human detection. In this work, there are three environmental sensors mounted at the front of the AGV, including an RGB-D camera, a laser scanner and a 3D optical sensor. The AGV platform is shown in Fig. 2. The laser scanner is used for the Lidar SLAM[4][5] of the AGV localization and also for avoiding unexpected obstacles. The RGB-D camera is installed at a height of 53 cm, it is responsible for human detection and AprilTag detection [6][7]. The 3D optical sensor is used for obstacle avoidance. In the current design, the AGV can autonomously navigate to the target points or the workstation without collision and ensure the safety of the surrounding human while docking to the workstation.

In order to improve the accuracy and stability of docking to the workstation, the AGV localization include two localization modules, the Lidar SLAM localization and AprilTag localization. The Lidar SLAM localization can locate the robot in the map and the AprilTag can locate the relative position between the robot and the workstation. The robot uses AprilTag localization while docking to the workstation and the Lidar SLAM localization [8] for path planning and navigation control. An L1 adaptive controller [9] was adopted to track the path point in order to reach the target docking station.

This wok was partly supported by the Ministry of Science and Technology of Taiwan R.O.C. under grant MOST 108-2221-E-009-128-MY2 and Syntec Technology.

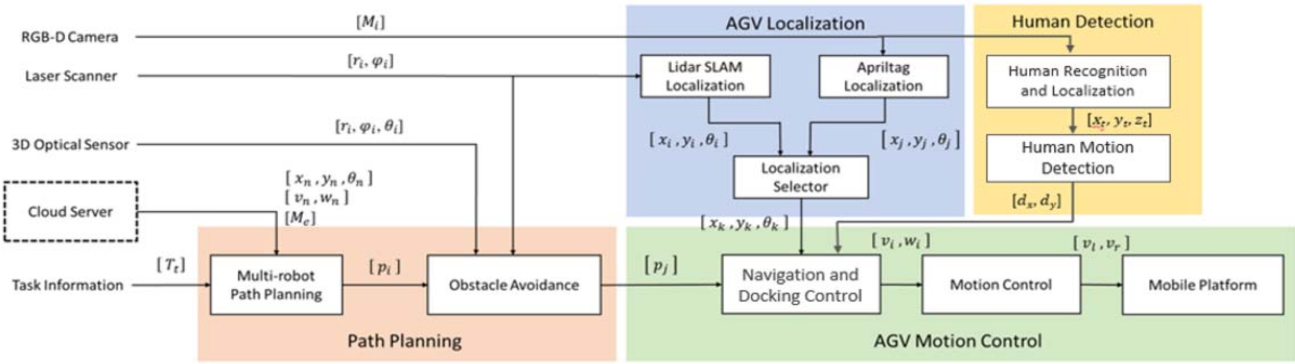


Fig. 1. The proposed system architecture for the Intelligent AGVs autonomous docking and human-robot collaboration

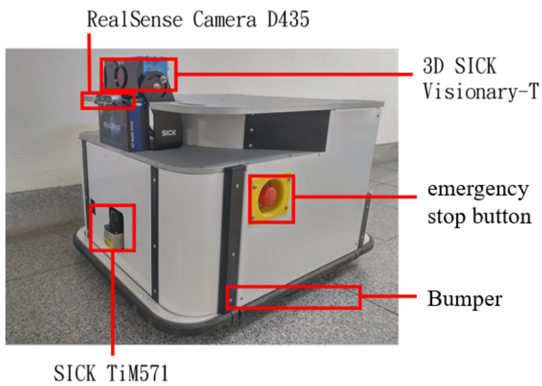


Fig. 2. The Intelligent AGV

The path planning module is responsible for planning a collision-free path from the robot position to the target points[8]. The Extended-ORCA algorithm developed in the previous work[10] was employed for multi-robot collision-free path generation for each robot to avoid mutual interference. For unexpected obstacles in the environment, a 2D obstacle map layer and a re-plan scheme are used for generating a collision-free path. In this work the A* algorithm[11] was adopted for path planning. In the multi-AGV system, each robot can share the same workspace with other robots and navigate to their target points safely without collision [10].

Human detection module can recognize the specific person and locate the relative position between the human and the robot. With human detection, the robot can autonomously move to the person who sends the request to call for help. Further, by using the human localization information, the AGV robot can also be aware of the people around it. A suitable motion will be planned to avoid collision with the person.

The docking workstation control module allows the AGV to plan an appropriate motion in a human-robot collaboration situation. If there are people near the docking position, the AGV must re-plan its motion based on human motion information. In this design, the AGV can prevent from collision and ensure safety while docking to the station.

III. Proposed Human-Motion Detection Method

In order for a smart AGV system to be able to perform human-robot collaboration, it must first know the human position in the environment. Fig. 3 depicts the proposed human-motion detection scheme. Firstly, after a human detection unit, a specific human recognition and localization module will recognize and locate a specific person. Then the motion information of the interested person will be obtained based on the current AGV location and the human location.

An RGB-D camera is employed to capture images for human detection and recognition. The effected range for human detection is around 7m. The MobileNetv2-SSDLite deep learning framework [12] was adopted to detect people in the image. The lightweight structure can fast select the position of all people in the image on the embedded platform, and provide for the next person identification unit to use, as shown in Fig. 4.

For the person identification, the Human Re-ID method was adopted for fast determining whether there is a specific person in the image. Fig. 5 illustrates the architecture of Human Re-ID, where the photo of the person to be recognized is represented as a query. And we compare the person framed by the detection network with the image of the query person, and calculate their similarity. Fig. 6 shows the person with the highest probability

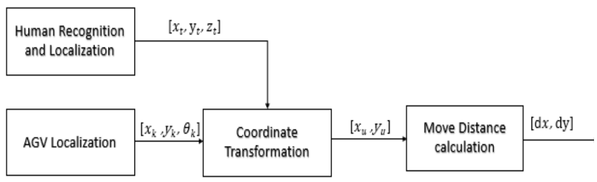


Fig. 3. The human motion detection system.

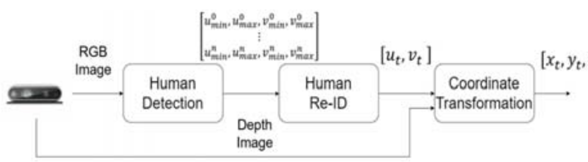


Fig. 4. Block diagram of human recognition and localization.

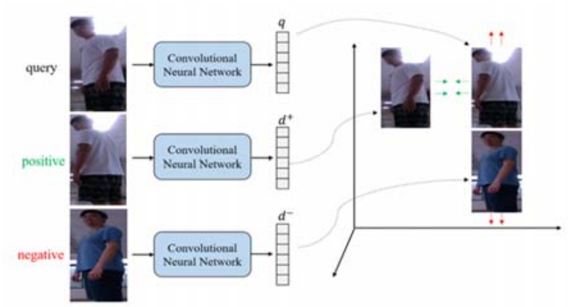


Fig. 5. The human identification learning network Human Re-ID.

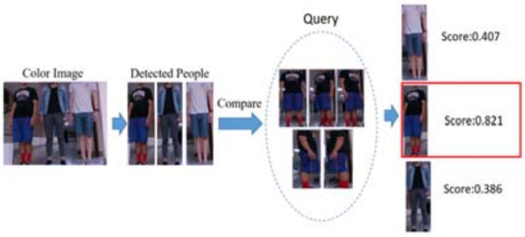


Fig. 6. The processing of human identification

and those greater than a certain threshold are regarded as the target persons.

A. Human Localization and Coordinate Transformation

Fig. 7 shows the result of recognition through the deep learning network Human Re-ID is framed with a bounding box, and the center of the box is set as the position of the person on the screen .

Based-on the pinhole model, the human location can be obtained,

$$u_t = \frac{u_l + u_r}{2} \tag{1}$$

$$v_t = v_l + \frac{v_r - v_l}{b} \tag{2}$$

where u_l, v_l, u_r, v_r are the coordinates of user's bounding box at upper left and lower right corner respectively, u_t, v_t are the center of user, and b is a proportional constant, which is used to determine the vertical position of the user in the pixel plane. Then the distance x_P, height h_P, and orientation θ, relationship between the person relative to the center of the camera are obtained as follows:

$$x_P = \frac{(u_t - c_x)z_c}{f_x} \tag{3}$$

$$h_P = \frac{(v_t - c_y)z_c}{f_y} \tag{4}$$

$$\theta = tan^{-1}(x_p/z_c) \tag{5}$$

where z_c is RealSense D435 camera in-depth information, and f_x , f_y represent the focal lengths of the camera

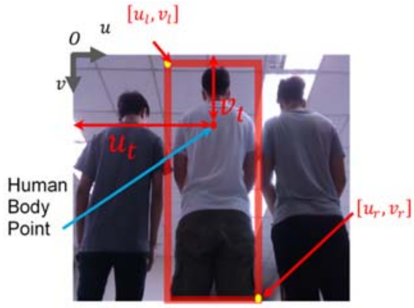


Fig. 7. Human position in the pixel plane

The coordinates of the person relative to the center of the camera are converted by pinhole model. Then the distance in the x, y, and z directions of the person from the center of the camera can be obtained as follows:

$$\begin{cases} x_t = z_c \sin\theta \\ y_t = z_c \cos\theta \\ z_t = h_p \end{cases} \tag{6}$$

where x_t , y_t , z_t represent the coordinates of the person's position from the center of the camera.

B. Human Motion Detection

By knowing of the person's position from the center of the camera. and the location of AGV, we can calculate the distance which the people move. Through the results of person recognition and positioning, the relationship between the person and the robot's camera is obtained, and the position information of the person on the map can be obtained through the coordinate transformation matrix of the robot's own position as follows:

$$\begin{bmatrix} x_u \\ y_u \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} cos\theta_k & -sin\theta_k & 0 & x_k \\ sin\theta_k & cos\theta_k & 0 & y_k \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ z_t \\ 1 \end{bmatrix} \tag{7}$$

where x_u, $y_u$represents the location of the person on the map, x_k, y_k, θ_k represents the position and rotation angle of the AGV on the map.

After obtaining the position of the human on the map, we can determine whether the person is moving by the person's position in the environment as follows:

$$\begin{cases} dx = x_u' - x_u \\ dy = y_u' - y_u \end{cases} \tag{8}$$

where dx, dy represent human movement distance of $X, Y\ axis$ per unit time, and x_u', y_u' is the location of the person on the map. The distance of the person movement in the sampled period is given by $\sqrt{dx^2 + dy^2}$.

IV. Docking Control With Human-Robot Collablration

In the current implementation, the error Lidar SLAM is about 10 cm, which is not sufficient for the AGV docking to perform high accuracy tasks. An AprilTag was used to make the robot perform a second localization calibration during docking to the workstation [6][7]. The RGB-D Camera Realsense D435 on the AGV is used to detect AprilTag, and to calculate its posture information. The position error and orientation error of the AprilTag docking is 2cm and 2.8 degree for the current system.

A. *Docking with Human Detection*

During the AGV docking, if a person is near the AGV docking point of the workstation, then the AGV need to handle the situation. The AGV determines whether the human is crossing the workstation or working at the workstation, thereby adjusting the speed of the AGV when docking to the station. The system architecture is shown in Fig. 8.

The control method of the AGV's linear velocity v_r and angular velocity w_r in the docking state are determined as follows:

$$v_r = \begin{cases} V_p \, , num = 0 \\ \frac{V_p}{3}, num > 0 \; and \; \sqrt{dx^2 + dy^2} \geq D_m \\ 0 \, , num > 0 \; and \; \sqrt{dx^2 + dy^2} < D_m \end{cases} \tag{9}$$

$$w_r = \begin{cases} w_{p \, ,num = 0} \\ 0_{\, ,num > 0} \end{cases} \tag{10}$$

where num is number of humans in the field of view. The AGV will first keeps the linear velocity, V_p, and the angular velocity, w_p, to perform docking to the station as given by the path planning module. If there is a person passing by during the docking task, the robot will calculate the distance of the person's movement $\sqrt{dx^2 + dy^2}$, judge whether the person is stop or moving by the constant D_m, and adjust the speed according to the formula.

B. *Docking to a Specific Human Location*

When the AGV is performing a navigation and docking task, if it receives a call command sent by a person, it will start to identify the calling person. If it recognizes the specific person, it will locate the person, reach to the position next to the person, and ready to receive the assigned task. If the AGV encounter other person, it will slow down and avoid the person to prevent a collision. The system architecture is shown in the Fig. 9.

The AGV uses Lidar SLAM localization and path planning algorithm which we described in Section II, to perform the navigation and docking task normally, when the human sends a call command, the AGV will stop the navigation mode it is performing and starts to find the specific person who is calling the AGV. The AGV uses the RGB-D camera in the front and uses MobileNetv2-SSDLite network to detect the people in sight, then it uses Human Re-ID neural network to recognize if the person.

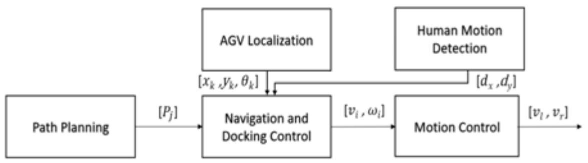


Fig. 8. Block diagram of docking with human detection

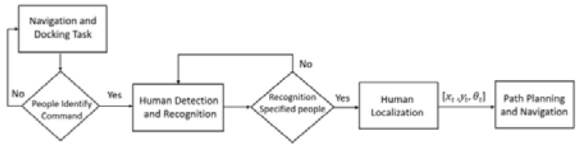


Fig. 9. Flow chart of docking to a specific human location

After detected and recognized the specific person who called for a help, the system can locate that human by the human localization algorithm as described in section III. After the AGV knows the location of that specific human, it will update the target point of the navigation module and uses its path planning algorithm to navigate to specific human, it will stay 30 cm next to the person and turn to the front, so that people can pick and place items easily for a human-robot collaboration task. If the human disappears from the AGV's vision, AGV will continue to reach the previous recorded point. On the way to the target, the obstacle avoidance algorithm is constantly operating, allowing the AGV to avoid obstacles and other humans so that it can safely navigate to that specific human.

V. Experimental Results

We implemented human motion detection method and integrated with the autonomous docking, to verify the proposed human-robot collaboration system on the AGV system. In the experiments, the local path planner is implemented by using open source move base [14]. The LiDAR SLAM and robot localization AMCL algorithm are implemented by open source [15][16]. Nvidia Jetson TX2 is employed on the AGV as the embedded platform. The local path planning and obstacle avoidance were integrated with a cloud computing platform to obtain the new path. The human motion information is obtained in real time to achieve human-robot collaboration. In the current system, the human recognition is excluded 15 frames per second(FPS), which is used for the human motion calculation. Two interesting experiments are presented to verify the proposed method autonomous docking and human-robot collaboration.

A. *Autonomous Docking*

In this experiment, the AGV encountered a moving person during navigation and docking. By using the proposed human motion detection and AGV control method described in Section III and IV, it slowed down and waited for the person to pass. During the process of docking at the workstation, if a person stopped to work at the workstation, the robot stopped and waited for the person to leave before docking to the workstation. The linear velocity v_r and angular velocity w_r was assigned based on the number of humans and the distance $\sqrt{dx^2 + dy^2}$ which human moved.

Fig. 10 shows the snapshots of the experiment of AGV navigation while encountering human passing. Fig.10(a) to Fig.10(d) show that when AGV encountered a person crossing while moving to the target location, the onboard camera detected the person and slowed down to wait for the person to pass. Fig.10(e) and Fig.10(f) show that after the person left, the AGV adjusted to the original speed to finish the work.

Fig. 11 shows the snapshots of the experiment of autonomous docking while there is a human working at the docking station. Fig.11(a) to Fig.11(c) show that the AGV performed docking by the localization of AprilTag. Fig.11(d) to Fig.11(f) show that when there was a human working in front of the workstation, the AGV detected the human and stopped the docking operation temporarily and waited for the human to leave. Fig.11(g) to Fig.11(i) show that after the person left, the AGV resumed to continue the task of docking to the workstation.

B. *Docking to a Human Location*

In this experiment, we designed a scenario to verify the human-robot collaboration. The idea is that when the AGV is in the process of docking to the station, if it receives a call from a specific human, then it will stop the current task and start to find the person. Through the human identification and motion detection system described in Section III, the location of the specific person can be known. After finding that specific person, the AGV will move to him for collaboration, and will evade other human in the environment. Further, in order to verify the flexibility of the system, the position of the specific human changes in the second part of the experiment, the AGV can also accurately stop by his side for human-robot collaboration.

Fig. 12 shows the snapshots of the experiment. In the start of experiment, the AGV performed its docking task, There were two human in the environment, as shown in Fig.12(a). Fig.12(b) shows that the AGV received a call from the specific person, it then stopped the current task and started to find the person who called for the help. The AGV recognized and located that person successfully. Fig.12(c) to Fig.12(f) show that the AGV dodged other people who were not the specific one when navigating toward the specific human. As shown in Fig.12(g) to Fig.12(i), the AGV successfully navigated to the specific person who called for a help and docked next to him for human-robot collaboration.

Fig. 13 shows that, in the second part of the experiment, if two people swapped their positions, the AGV still recognized and located the specific human, and successfully docked next to him. Fig.13(a) and Fig.13(b) show that the AGV rotated itself and recognized the specific human. As shown in Fig.13(c) to Fig.13(f), the AGV successfully located the specific human and docked next to him for human-robot collaboration.

VI. Conclusions And Future Work

In this paper, we propose an intelligent AGV system, which can perform automatic docking to a station and human-robot collaboration. The experiments demonstrated that the AGV can automatically navigate to the target locations or workstations based on LiDAR SLAM and AprilTag. During the navigation process, the LiDAR and the RGB-D camera in front of the robot

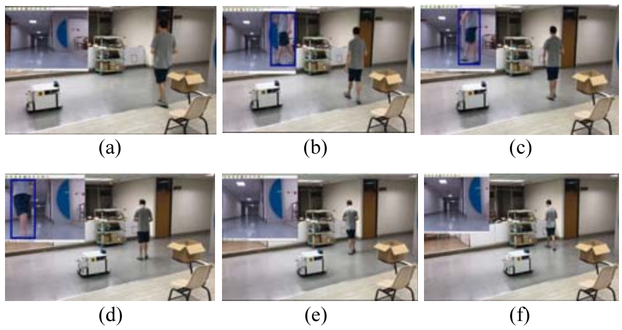


Fig. 10. Snapshots of the experiment of navigation while encountering human passing. (a)~(d) When the AGV encounters a person. (e)~(f) When the person passed by, the AGV resumed the navigation task.

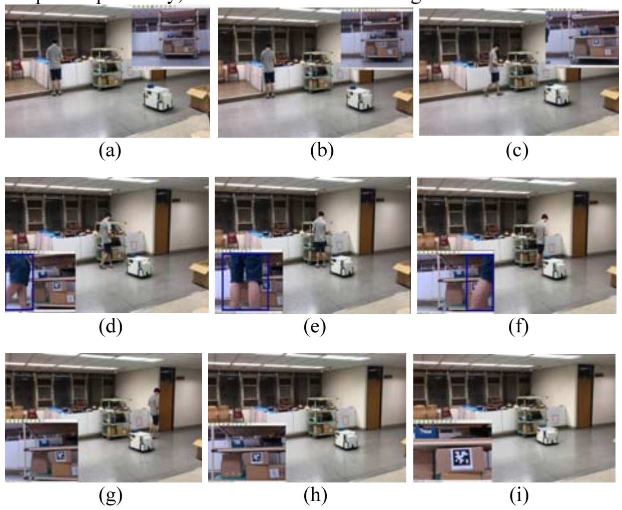


Fig. 11. Snapshots of the experiment of autonomous docking while there is a human stopped at the docking station. (a)~(c) AGV performed docking by the location of AprilTag. (d)~(f) The AGV detected the human and stop the docking test and wait for the human leave. (g)~(i) after the person left, the AGV continued to finish the docking task.

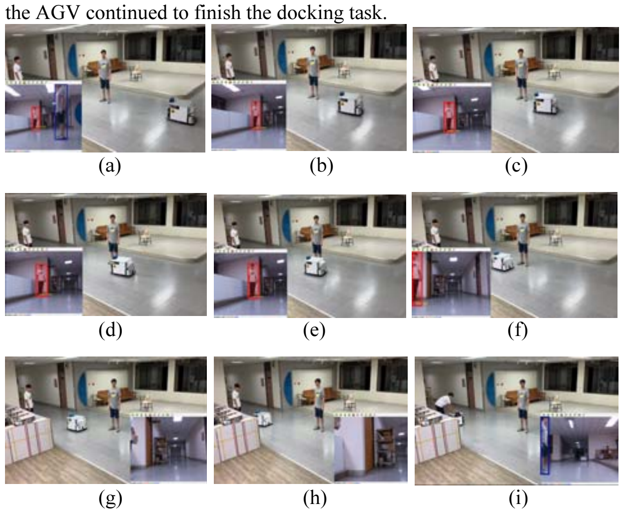
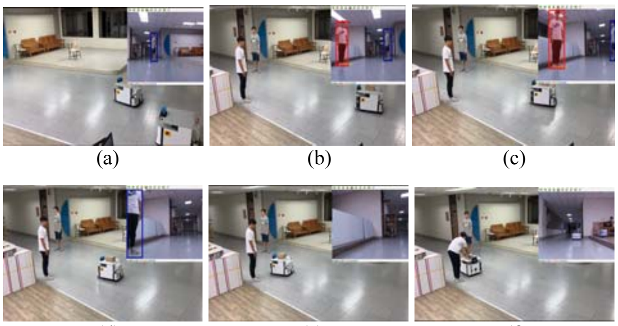


Fig. 12. Snapshots of the experiment of docking to a human location. (a) There were two humans in the environment. (b) The AGV received a call to help, it started to find the person who called for the help. (c)~(f)The AGV dodged other people who were not the specific one when navigating toward the specific human. (g)~(i) The AGV successfully docked in front of the specific person.



(a) (b) (c)

(d) (e) (f)

Fig. 13. Snapshots of the experiment of the case that the specific human changed his position in the docking-to-a-human-location experiment. (a)~(b) The AGV rotated itself and recognized the specific human. (c)~(f) The AGV successfully located the specific human and docked next to him for human-robot collaboration.

continuously detect whether there are obstacles or human nearby. In our purposed human motion detection system, the camera in front of the robot instantly identify people in the robot's field of view and calculated their positions and motion in real time. So the AGV can avoid collisions and ensure the safety of surrounding people. When people need for a help, they can also call the AGV to dock next to him for human-robot collaboration. In the future, a robotic arm will be setup on the AGV as well as at the workstation to further improve the capacity of the AGV system.

ACKNOWLEDGMENT

The authors are grateful for Pei-Chun Lu, and Zhen-Yu Wang for their assistance in the experiments. The support of Syntec Technology for this project is greatly appreciated.

REFERENCES

[1] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *Proc. of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1701-1708.

[2] F. Schroff, D. Kalenichenko and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR*), Boston, USA, 2015, pp. 815-823.

[3] J. Almazan, B. Gajic, N. Murray and D. Larlus, "Re-ID Done Right: Towards Good Practices for Person Re-identification," *arXiv preprint arXiv:1801.05339,* 2018.

[4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99-110, June 2006.

[5] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): part II," *IEEE Robotics & Automation Magazine*, vol. 13, no. 3, pp. 108-117, Sept. 2006.

[6] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proc. of 2011 IEEE International Conference on Robotics and Automation*, Shanghai, China, 2011, pp. 3400-3407.

[7] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proc. of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, Korea, 2016, pp. 4193-4198.

[8] K. Song, Y. Chiu, L. Kang, S. Song, C. Yang, P. Lu and S. Ou, "Navigation Control Design of Mobile Robot by Integrating Obstacle Avoidance and LiDAR SLAM," in *Proc. of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, 2018, pp. 1833–1838.

[9] T. Fukao, H. Nakagawa and N. Adachi, "Adaptive tracking control of a nonholonomic mobile robot,"in *IEEE transactions on Robotics and Automation*, vol.16, no. 5, pp. 609-615, Oct. 2000.

[10] K. T. Song and Y. X. Sun, "Coordinating Multiple Mobile Robots for Obstacle Avoidance Using Cloud Computing," to appear in *Asian Journal of Control*, 2020.

[11] C. Wang, L. Wang, J. Qin, Z. Wu, L. Duan, Z. Li and Z. Lu, "Path planning of automated guided vehicles based on improved A-Star algorithm," in *Proc. of 2015 IEEE International Conference on Information and Automation*, Lijiang, China, 2015, pp. 2071-2076.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,* Salt Lake City, UT, USA, 2018, pp. 4510-4520

[13] A. Krizhevsky, I. Sutskever, E.G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing System,* pp. 1097-1105, 2012.

[14] "move_base – ros wiki," wiki.ros.org/move_base [accessed : August 1, 2020]

[15] " Cartographer ROS Integration," http://googlecartographerros.readthedocs.io[accessed : August 1, 2020]

[16] "AMCL – ros wiki," wiki.ros.org/amcl [accessed : August 1, 2020]