A Multimodal Person-following System for Telepresence Applications

Wee Ching Pang*
BeingThere Centre
Institute for Media Innovation
Nanyang Technological University

Gerald Seet† and Xiling Yao‡
Robotics Research Centre
School of Mechanical and Aerosapce Engineering
Nanyang Technological University

Abstract

This paper presents the design and implementation of a multimodal person-following system for a mobile telepresence robot. A color histogram matching and position matching algorithm was developed for a person-recognition function using Kinect sensors. Robot motion was controlled by adjusting its velocity according to the humans position in relation to the robot. The robot was able to follow the targeted person in various person-following modes, such as the back-following mode, the side-by-side accompaniment mode as well as the front-guiding mode. An obstacle avoidance function was also implemented using the virtual potential field algorithm.

CR Categories: I.2.9 [Robotics]: Autonomous vehicles—Sensors H.4.3 [Communications Applications]: Computer conferencing, teleconferencing and videoconferencing—;

Keywords: Telepresence robot, multimodal person-following, Kinect, back-following, side-by-side accompaniment, escorting, guiding, autonomous

Links: DL PDF

1 Introduction

With the prevalence of the Internet and the advances of technology, there is a continuing desire to replace face-to-face interaction with one that can make do without the need for time consuming international travel. This is evident with the increased usage of video-mediated communications such as video conferences within boardroom and mobile video chatting tools such as Microsoft Skype, Apple FaceTime and Google Talk. Although such communication via video screen provides rich interaction experience, it lacks a number of spatial cues and mobility.

In recent past, there has been an emergence of new communication means where a robot is used as the communication mediums by embodying video conferencing on wheels. This has allowed the robots user (inhabitor) to see, hear and move together with the person with whom he or she is communicating, while this person who is interacting with the robot (interactant) is situated miles away with the robot. Such robot-mediated communication tools are commonly known as telepresence robots although some may call them remote presence systems [Willow Garage 2011], virtual presence systems [Anybots Inc. 2013], embodied social proxy [Venolia et al. 2010] or robotic avatars [Lincoln et al. 2009] [Seet et al. 2012b] [Seet et al. 2012a].

A telepresence robot called MAVEN [Seet et al. 2012b] has been developed for this work. MAVEN is a synergistic fusion of intelligent mobile robot technology enabled with telepresence and video conferencing functionality [Seet et al. 2012a]. It allows the inhabitor to establish his or her presence with the use of a 2D graphical display as seen in Figure 1(a), or 3D physical display, as depicted in Figure 1(b) and (c) (in this setup, the inhabitor's face and head motions are projected on a 3DOF robotic head [Pang et al. 2012]) or a 2D transparent screen as shown in Figure 1(d). The audio, video and data link of the teleconferencing module would permit the transmission of speech, images and motion commands from the inhabitors site to the place where MAVEN resides.

(a) (b) (c) (d)

Figure 1: *MAVEN telepresence robots with (a) 2D graphical display, (b) 3D physical display, (c) 3D physical avatar and (d) 2D avatar shown on a transparent display*

Unlike the conventional static teleconference communication, the addition of mobility on such robot-mediated communication can increase the cognitive workload of the inhabitor acutely. It can subsequently distract the inhabitor from the task of interaction. In an effort to reduce the cognitive load resulting from piloting the robot, the robot has to be equipped with the ability for fully or semi-autonomous motion navigation. With such capabilities, the robot would be capable of behaviors like obstacle avoidance, collision prevention, human following as well as other human-like motion behaviors.

The inhabitor controlling MAVEN is expected to control the telepresence robot to move together with a person while simultaneously having a conversation. MAVEN is thus equipped with a multimodal person-following system that provides the functionality of "walking with a person" autonomously. This system is expected to alleviate the workload of robot control. In this work, three different person-following modes have been identified, in which a telepresence robot

*e-mail: weeching@ntu.edu.sg
†e-mail: mglseet@ntu.edu.sg
‡e-mail: yaox0006@e.ntu.edu.sg

can move together with a person. These modes are illustrated in Figure 2.

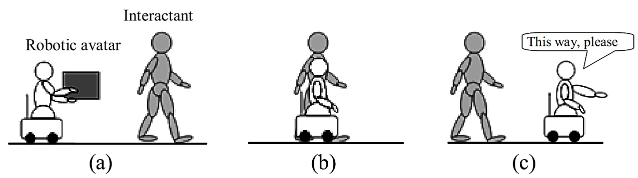

Figure 2: *Different modes of person-following - (a) back following, (b) side-by-side accompaniement and (c) front guiding*

In the back-following mode, the robotic avatar will be moving behind the person whom it is following. For instance, an inhabitor, who is using a telepresence robot to visit a museum, can choose to use the robot to autonomously follow a human guide. This mode can also be used to enable the robotic avatar to follow an interactant and provide assistance as shown in Figure 2(a).

In the side-by-side accompaniment mode, as seen in Figure 2(b), the telepresence robot will be moving at the side of the interactant whom it is tracking. This is typical of a scenario where two people are having a conversation while walking side by side. In this case, one of the two people in the scenario is replaced with the telepresence robot that is controlled by the remote inhabitor.

In the escorting or guiding mode, as illustrated in Figure 2(c), the robotic avatar will be moving in front of its interactant, such that the inhabitor can provide direction using the telepresence robot to guide the interactant. This mode is useful when the inhabitor is more familiar with the robot's environment than the interactant.

During the autonomous person-following task, the robot would try to maintain a desirable standoff distance from the interactant. The obstacle avoidance capability has been implemented within the system to ensure safe navigation in human environment. Furthermore, as the person-following system is implemented for a telepresence robot, it is essential for the robot to move at a speed that is similar to that of a real human.

The scope of this paper includes the design and development of the proposed multimodal person-following module for telepresence applications. Firstly, a relevant literature review on person-following algorithms as well as on social navigation methodologies is presented in Section 2. A brief description of MAVEN, the robotic platform that is used in this study, is in Section 3. In Section 4, the architecture of the multimodal person-following system is specified in detail. Lastly, the result is presented in Section 5.

2 Related Work

The task of person-following involves human detection and tracking, which uses the data acquired from a sensor. One method of performing this task is to detect and track human using a digital color (RGB) camera. Image processing techniques can be applied to the images acquired from the camera to identify blobs that signify a person. Color detection or color histograms [Jin et al. 2010] [Kwon et al. 2005] and feature recognition [Chen and Birchfield 2007][Yoshimi et al. 2006] are some of the other common techniques that have been considered.

The laser range finder is another widely used device for robots to observe the environment. Compared with camera vision data, laser data is more efficient and hence less processing is required [Fod et al. 2002]. The distance measurements of a laser range finder usually have high accuracy, and the data is not sensitive to ambient noise like the changing lighting conditions. Therefore, there are many laser-based human detection and tracking studies [Topp and Christensen 2005][Arras et al. 2007][Gockley et al. 2007], which use techniques that process the ranging data to identify the signature of a persons legs. However, in some indoor applications, chairs and tables can be falsely detected as human legs due to the furniture having similar patterns as a humans legs.

Human detection can also be achieved with the use of depth camera [Loper et al. 2009], as well as with the new and inexpensive Kinect RGB-D camera [Luber et al. 2011]. In this paper, the Kinect RGB-D camera is used to detect and track the person to be followed. Although the field of view of the Kinect sensor is small, it provides sufficient information, such as depth values, audio data, skeletal mapping as well as a color image, to perform human detection.

Some of the related work demonstrates the various mode of person following, which include the back-following mode (such as [Gockley et al. 2007][Loper et al. 2009] and [Cosgun et al. 2013]), the side-by-side accompaniment mode [Prassler et al. 2002][Ohya and Munekata 2002] as well as the front guiding mode [Montemerlo et al. 2002][Pacchierotti et al. 2006][Burgard et al. 1998]. However, there does not appear to be existing work on combining these following modes into one system. Moreover, telepresence robots and humans will co-exist in the same space hence it is important that the robot would be able to move in a manner that is acceptable by the humans around it. There are many works on social navigation that have enabled mobile robots to move from one point to another in the presence of human [Topp and Christensen 2005][Gockley et al. 2007][Burgard et al. 1998], using proxemics [Hall 1969][Kirby et al. 2009]. However, the navigation system for a telepresence robot should be different from that of a social robot because, unlike a social robot, there is an additional intelligence from the human inhabitor behind a telepresence robot.

3 Hardware Configuration

The multimodal person-following system is implemented on MAVEN-II, which is one of the experimental telepresence robots as shown in Figure 1(b) and (c). MAVEN-II is a holonomic robot with 4 mecanum wheels. It has an on-board computer for controlling the drive motors. A Fedora operating system has been installed on the robots embedded computer. For this experiment, the maximum forward and lateral speed of the robot has been limited to 0.6 m/s while the rotational speed has been limited to 0.9 rad/s.

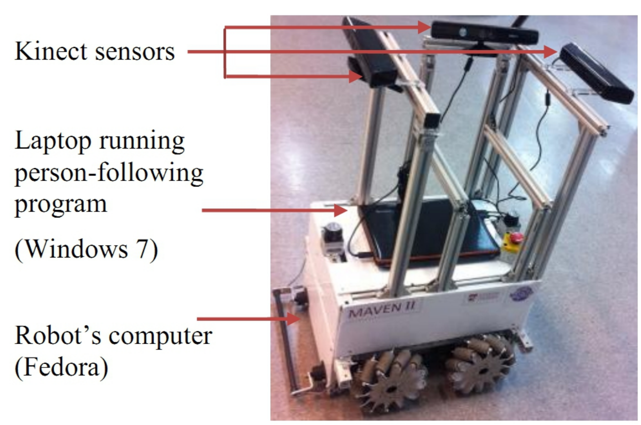

Figure 3: *Hardware configuration of MAVEN-II*

The multimodal person-following system has been implemented on an additional laptop computer that runs the Microsoft Kinect SDK.

Three Kinect sensors have been mounted on MAVEN-II, as seen in Figure 3; one of them is forward-looking and faces the front of the robot while the other faces the side of the robot. The third Kinect is mounted at the rear of the robot. These Kinect sensors are used to track the selected interactant and they were connected to the laptop. This laptop is responsible for acquiring data from the Kinect sensors, running the person-following algorithm, and sending velocity and control commands to the embedded computer which in turn controls the robot's movement.

4 The Multimodal Person-following System

The person-following system comprises of three main components human detection and tracking, person-following behavior and obstacle avoidance.

4.1 Human detection and tracking using Kinect and Kinect SDK

The Kinect SDK v1.5 provided by Microsoft is a set of tools and APIs that are used to implement applications, which use a Kinect sensor. The Natural User Interface module of the Kinect SDK provides the functionality of accessing the RGB color image data and depth data from the Kinect sensor, and it can be used to detect and track people within the field of view of the Kinect sensor.

The Kinect sensor has the capability to detect up to six people in its field of view and it is able to obtain detailed skeletal information, such as the positions and orientations of joints, for a maximum of 2 people. The Kinect sensor can detect and track people who stand between 0.8 m and 4.0 m from the front of the sensor. Each successfully tracked human can have one of the two tracking states, that is, the "position only" state or the "active user tracking" state.

In the "position only" state, only the real-world position (in meters) of the person can be obtained and no other information is available. In the "active user tracking" state, both of the centroid position and the skeletal data, which includes positions and orientations of various joints, are tracked. An ID will be randomly assigned to each detected person and the target can be chosen by selecting the ID of that detected humans.

Figure 4 shows a depth image acquired from the Kinect. In that figure, two people are detected within the sensors field of view. The ID as well as the centroid position of the depth image from each person is displayed. The centroid is represented by a blue circle. The inhabitor can initiate the person-following behavior by clicking on the centroid of the targeted person. Once a person has been selected as the target, the blue circle will turn into a red one to indicate that active tracking is in progress.

In addition to person-tracking, a person-recognition function is required to keep track of the person even when that person is temporarily out of view. An algorithm, which is based on the calculated Euclidean distance between the positions and a color histogram matching technique, has been included to identify and recognize the person.

When a new person is detected by the Kinect sensor after the previously targeted person is temporarily occluded or out of the scene, the newly detected person's current position and HSV color histogram will be compared against the stored position and histogram data. With this comparison, the person with the closest matching position and histogram data will be regarded as the previously tracked person, and the system will automatically resume active tracking.

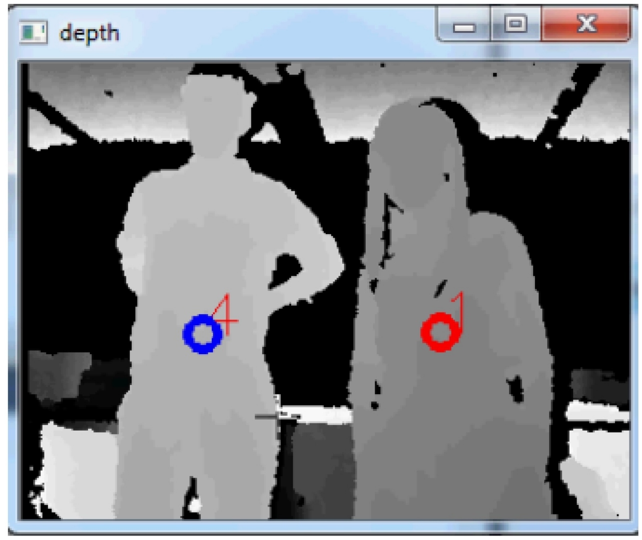

Figure 4: *Depth image showing User ID and centroid of the detected human's blobs*

The position matching measures the Euclidean distance between positions (d_{pos}) as shown in Equation 1, where C_T is the centroid position of the previously tracked person P_T and C_D is the centroid position of the detected person P_D.

$$d_{pos}(P_T, P_D) = \sqrt{\begin{matrix}(C_{T,x} - C_{D,x})^2 + (C_{T,y} - C_{D,y})^2 + \\ (C_{T,z} - C_{D,z})^2\end{matrix}} \quad (1)$$

The histogram matching techniques compares the histogram of the previously tracked person, H_T and the histogram of each detected person, H_D with four types of histogram comparison methods [Bradski and Kaehler 2008]: Correlation, Chi-square, Intersection and Bhattacharyya. The final histogram matching distance result is an arithmetic combination of all four histogram distances obtained from these methods as shown in Equation 2.

$$\begin{aligned} d_{hist}(H_T, H_D) =& (1 - d_{correlation}(H_T, H_D)) \\ &+ d_{chisquare}(H_T, H_D) \\ &+ (1 - d_{intersect}(H_T, H_D)) \\ &+ d_{Bhattacharyya}(H_T, H_D) \end{aligned} \quad (2)$$

Subsequently, a score S will be computed for each detected people based on their position and histogram matching result, as shown in Equation 3. The detected person with the lowest score is most likely to be the previously tracked person, and the person-following behavior will resume.

$$S = \omega\, d_{pos}(P_T, P_D) + \frac{(1 - \omega)\, d_{hist}(H_T, H_D)}{(max(\forall D : d_{hist}(H_T, H_D))} \quad (3)$$

4.2 Person-following Behavior

Three person-following modes for the telepresence robot have been implemented. They are the back-following mode, the side-by-side accompaniment mode and front-guiding mode. At each iteration, the robot-human linear and angular differences are measured and new velocities commands are computed to move the robot accordingly. The person-following process is essentially a proportional feedback control loop which measures the error between the current robot position and the desired robot position with respect to the targeted person. The error is then used to calculate the velocity commands of the robot. The control loop is depicted in the control diagram in Figure 5.

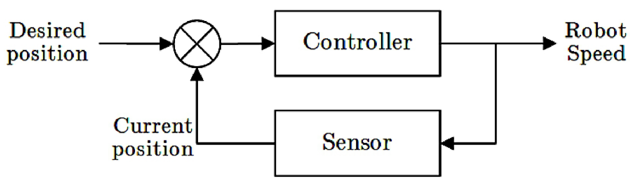

Figure 5: *Feedback control loop for robot speed calculation*

The controller block represents the calculation formulae for the robots forward, lateral and rotational velocities. Each person-following mode will have its own controller. The implementation of these controllers is described in the subsequent subsections.

4.2.1 Person-following mode 1: Back Following

In back-following mode, the followed interactant stays in front of the robotic avatar while walking around. The robot is able to move, by translating and rotating, with the interactant while trying to maintain a predefine 1.5 m distance to that interactant. The value 1.5 m was chosen because it is the common "social distance" between two people during social interactions [Hall 1969]. The robot will also try to rotate itself in order to keep the interactant at the center of the Kinect sensor's field of view.

The desired distance and direction of the targeted person relative to the robot is noted as the "neutral distance" and "neutral direction" respectively, which will be used as the reference for robot speed calculation. The forward translation velocity (tv), is directly proportional to the relative range between the current distance of the human from the robot and the neutral distance, as shown in Figure 6. Similarly, the rotational velocity (rv) is directly proportional the angular difference between the predefined neutral direction and the current direction of the targeted person, as shown in Figure 7. In this mode, the strafe (or sidestepping) velocity (sv) is not used to generate the motion for person-following.

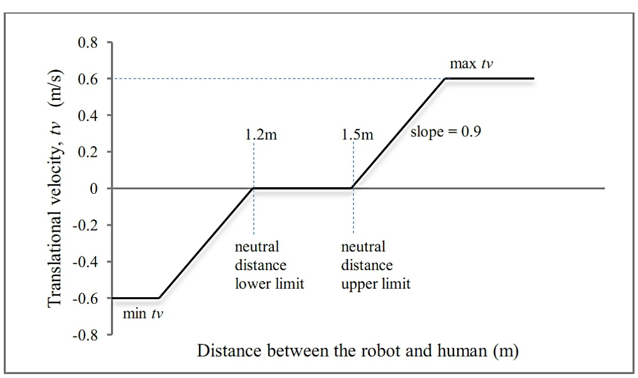

Figure 6: *Distance and translational velocity relationship*

4.2.2 Person-following mode 2: Side-by-side Accompaniment

In the side-by-side accompaniment mode, the Kinect sensor is mounted at the side of the robot such that the Z_0 axis of the sensor points to the left-hand-side direction of the robot. In this manner, the sensor is able to track the interactant while the interactant

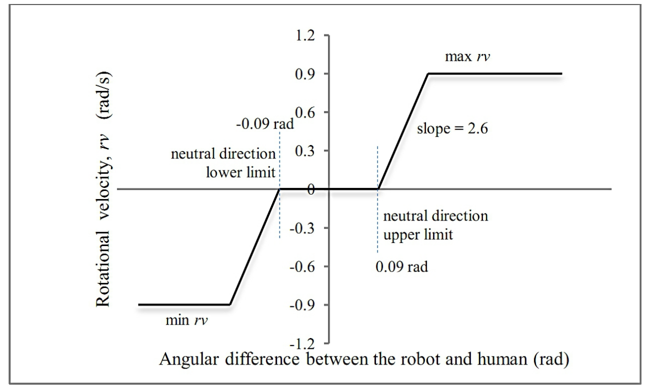

Figure 7: *Relationship between the angular difference and rotational speed*

walks besides the robotic avatar on its left-hand-side. The setup is as shown in Figure 8(a), where the relative position between the robot and the human is described. Three quantities, including the robot-human distance, the rotation angle as well as the offset, will be used in the computation of velocities commands during side-by-side accompaniment mode. The robot-human distance and offset distance can be obtained directly from the Kinect SDK.

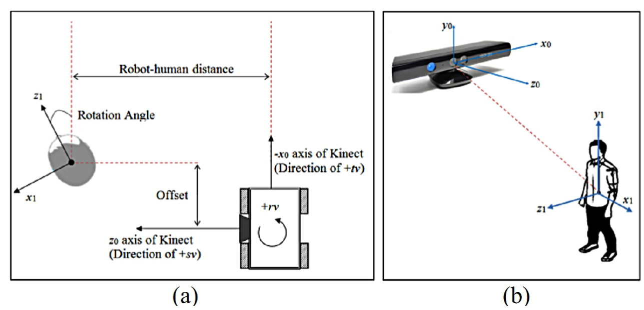

Figure 8: *(a) Relative robot-human positions in Side-by-side accompaniment (b) Human frame and Kinect frame*

Figure 8(b) illustrates the coordination system of the interactant's human frame with respect to the Kinect frame. The Kinect frame is X_0-Y_0-Z_0 and the human frame is X_1-Y_1-Z_1. The rotation matrix from X_0-Y_0-Z_0 to X_1-Y_1-Z_1 can be obtained with the Microsoft Kinect SDK, and it is of the following format

$$R_0{}^1 = \begin{bmatrix} M_{11} & M_{12} & M_{13} & 0 \\ M_{21} & M_{22} & M_{23} & 0 \\ M_{31} & M_{32} & M_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

The 3x1 vectors $\begin{pmatrix} M_{11} \\ M_{21} \\ M_{31} \end{pmatrix}$, $\begin{pmatrix} M_{12} \\ M_{22} \\ M_{32} \end{pmatrix}$ and $\begin{pmatrix} M_{13} \\ M_{23} \\ M_{33} \end{pmatrix}$ are the unit vectors of the axes X_1, Y_1 and Z_1 respectively and they represent the orientation of X_1, Y_1 and Z_1 with respect to the Kinect frame. In the side-by-side accompaniment mode, the direction at which the interactant is facing is that of Z_1. Therefore, the unit vector of Z_1 axis is used to calculate the rotation angle that the robot must rotate so that it faces the same direction as the interactant it is following, as shown in Equation 5.

$$\text{Rotation Angle} = \tan^{-1}\left(\frac{M_{33}}{-M_{13}}\right) \tag{5}$$

The velocities commands that produce the motion of the robot during the side-by-side accompaniment include translational velocity (tv), strafe velocity (sv) and rotational velocity (rv). The offset distance will determine the translational velocity and the rotation angle will influence the rotational velocity, while the strafe velocity is dependent on the robot-human distance. The tv-offset, rv-rotation angle and sv-distance relationships are directly proportional and they are depicted in Figure 9, Figure 10 and Figure 11 respectively.

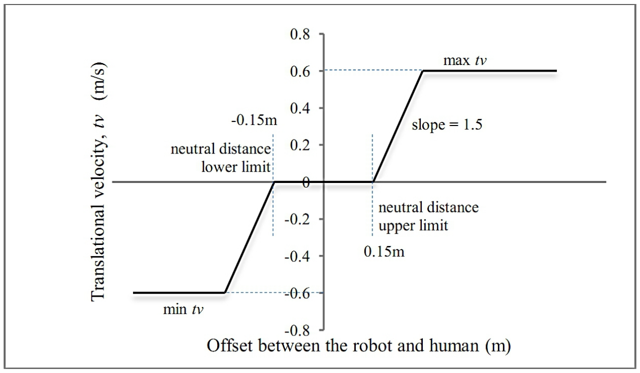

Figure 9: *Offset and translational velocity relationship*

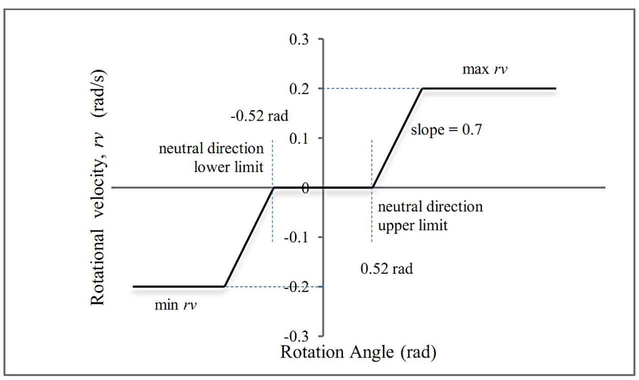

Figure 10: *Relationship between the rotation angle and rotational speed*

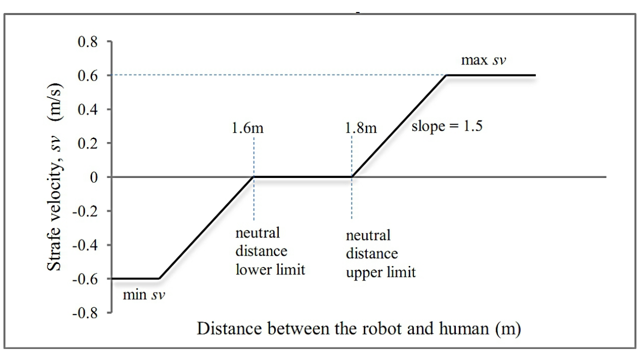

Figure 11: *Distance and strafe velocity relationship*

4.2.3 Person-following mode 3: Front Guiding

In the front guiding mode, the telepresence robot will be moving in front of the interactant and its task is to escort the interactant around. Therefore, unlike the back-following and the side-by-side accompaniment modes, the robot in this mode will move to a designated goal position, which is independent of the interactant's movements.

Although the robot knows where to go, it still tracks the interactant so that it can adjust its velocity according to the interactant's movement. The robot will still try to maintain a pre-defined 1.5 m distance from the interactant. If the interactant lags behind and the distance between the interactant and robot is greater than 1.5 m, the robot will slow down to accommodate the interactant. On the other hand, if the interactant increases his or her walking speed, the robot will move more quickly towards the goal position. In this manner, the forward translation velocity (tv), is directly proportional to the relative range between the current distance of the person from the robot as shown in Figure 12.

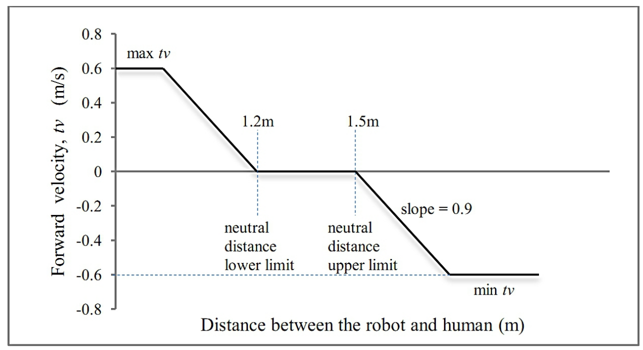

Figure 12: *Distance and translational velocity relationship*

4.3 Obstacle Avoidance

Currently, the obstacle avoidance capability has only been implemented for the front-guiding and back-following modes. The robots strafe velocity will be used to perform obstacle avoidance while the robot is in these modes. In this manner, the robot will avoid obstacles by "side-stepping".

The reactive obstacle avoidance capability has been implemented using the concept of virtual potential field (VPF) [Koren and Borenstein 1991]. The virtual repel force exerted by the obstacle is inversely proportional to the distance (or distanced raised to a certain power) between the obstacle and the moving robot. That is, if the robot comes closer to the obstacle, the virtual repel force on it from the obstacle will be larger. The robot would then move away from the obstacle with a higher velocity.

While the Kinect sensor is used for tracking the interactant, it is also used to observe obstacles between the robot and the interactant. The Kinect sensor is able to obtain the 3D position of all points on the depth image. The system would first calculate the current distance from the robot to the interactant. For each pixel on the Kinect sensor's depth image that corresponds to a distance less than that between the robot and the interactant, the system would consider that as belonging to an obstacle. A virtual repel force is then exerted on the robot by that point. After all obstacle points have been identified, the virtual repel forces contributed by these points are then summed up. Finally, the summation of virtual repel forces is divided by the total number of obstacle points, resulting in the average virtual repel force which would then be used to calculate the robot's velocity in order to avoid obstacles. The virtual repel force, F_{repel}, from each obstacle point is given by Equation 6, where K is a constant and D is the distance from the point to the robot.

$$F_{repel} = \frac{K}{D^3} \qquad (K = 0.6) \tag{6}$$

For small value of D, the resulting virtual repel force will be larger. In order for the robot to efficiently avoid obstacles during the front-following and the back-following modes, the robot's strafe velocity (sv) was calculated from the average virtual repel force and then transmitted to MAVEN-II. The translational velocity (tv) and rotational velocity (rv) were not affected by the existence of obstacles; they were determined only by the position of the followed person relative to the robot. When tv, sv and rv are combined, the resulting motion of the robot enables obstacle avoidance. At the same time, the robot would also be able to maintain a desirable distance and orientation towards the interactant. The virtual repel force is proportional to the incremental amount of transverse speed sv, as shown in Equation 7, where A is a predefined coefficient, and $\triangle sv$ represents the incremental sv value after each update of image and depth frame from Kinect sensor.

$$\triangle sv = A \times F_{repel} \tag{7}$$

The resulting transverse robot speed can be calculated using Equation 8, where A is set to 1 for simplicity and n is the current number of Kinect data frames, counting from the moment when an obstacle has been detected (when n=0).

$$sv = \triangle sv \times n = A \times F_{repel} \times n \tag{8}$$

The sv value will gradually increase with the increasing value of n, so that the obstacle avoidance movement will not be too abrupt. Since the frame update rate of Kinect is fast (minimum 9 Hz), the sv value can increase from zero up to its maximum speed (0.3m/s) within 1 second. This would allow the robot to quickly avoid nearby obstacles.

5 Results

The multimode person-following system has been implemented. Experiments have been carried out to test the robustness of the back-following behavior as well as the side-by-side accompaniment behavior. The trajectories of the robot and the human interactant in the back-following mode and the side-by-side accompaniment mode are as shown in Figure 13 and Figure 14 respectively.

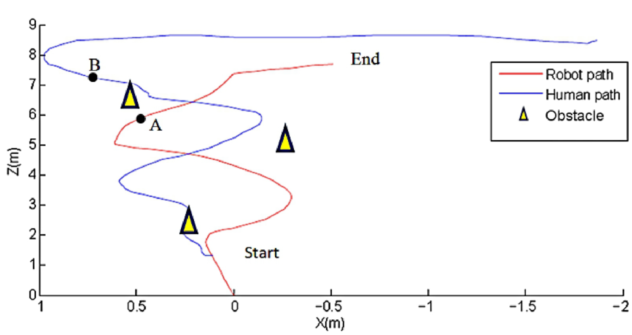

Figure 13: *Trajectories of the robot and human interactant in back-following mode*

As shown in Figure 13, the robot did not follow the interactant by moving on path as the human interactant. Instead, the robot adjusted its direction continuously to ensure that the followed person is always centered in the Kinect sensors field of view. This would reduce the distance travelled by the robot during the process of person-following. It also resembles the behavior of a real human when he or she is following another person. This is because a human would not typically attempt to follow the exact path taken by the leader.

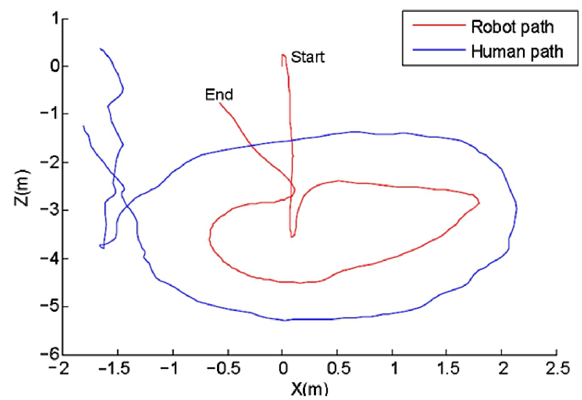

Figure 14: *Trajectories of the robot and human interactant in side-by-side accompaniment mode*

In Figure 14, the robot was able to perform side-by-side accompaniment by moving to stay at the side of the targeted person. However, the motion of the robot was not always stable due to the fluctuation and inaccurate calculation of the followed persons orientation.

6 Conclusion

In this project, a multimodal person-following system has been designed and implemented on a mobile robotic avatar, which can be used for telepresence applications. The person-following system includes three different person-following modes such as back-following, side-by-side accompaniment and front-guiding or escorting. Obstacle avoidance function has also been implemented for the back-following mode and front-guiding mode.

The main contribution of the paper is the novel way in which a system, which includes multiple modes of person-following behaviors with obstacle avoidance, has been implemented. This novelty stems from the use of Kinect sensors for tracking the interactant.

The experiment results show that the back-following mode was robust with small chance of losing track of the targeted person. The side-by-side accompaniment mode was less robust in comparison. This is because of the higher likelihood of losing track of target and the higher propensity for the robot to move in an unsteady manner when the robot was in the side-by-side accompaniment mode.

7 Acknowledgments

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Program.

References

ANYBOTS INC., 2013. Anybots virtual presence systems. it's you, anywhere... Available online at https://www.anybots.com/.

ARRAS, K. O., MOZOS, O., AND BURGARD, W. 2007. Using boosted features for the detection of people in 2d range data. *IEEE International Conference on Robotics and Automation (ICRA)*, 3402–3407.

BRADSKI, G., AND KAEHLER, A. 2008. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Incorporated.

BURGARD, W., CREMERS, A. B., FOX, D., HHNEL, D., LAKEMEYER, G., SCHULZ, D., STEINER, W., AND THRUN, S. 1998. The interactive museum tour-guide robot. *Proceedings of the National Conference on Artificial Intelligence*, 11–18.

CHEN, Z. C., AND BIRCHFIELD, S. T. 2007. Person following with a mobile robot using binocular feature-based tracking. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 815–820.

COSGUN, A., FLORENCIO, D. A., AND CHRISTENSEN, H. I. 2013. Autonomous person following for telepresence robots. *IEEE International Conference on Robotics and Automation (ICRA)*.

FOD, A., HOWARD, A., AND MATARIC, M. J. 2002. A laser-based people tracker. *IEEE International Conference on Robotics and Automation (ICRA) 3*, 3024–3029.

GOCKLEY, R., FORLIZZI, J., AND SIMMONS, R. 2007. Natural person-following behavior for social robots. *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, 17–24.

HALL, E. T. 1969. *The hidden dimension*. Anchor Books New York.

JIN, L., CHENG, J., AND HUANG, H. 2010. Human tracking in the complicated background by particle filter using color-histogram and hog. *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 1–4.

KIRBY, R., SIMMONS, R., AND FORLIZZI, J. 2009. Companion: a constraint-optimizing method for person-acceptable navigation. *The 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 607–612.

KOREN, Y., AND BORENSTEIN, J. 1991. Potential field methods and their inherent limitations for mobile robot navigation. *IEEE International Conference on Robotics and Automation (ICRA)*, 1398–1404.

KWON, H., YOON, Y., PARK, J. B., AND KAK, A. C. 2005. Person tracking with a mobile robot using two uncalibrated independently moving cameras. *IEEE International Conference on Robotics and Automation (ICRA)*, 2877–2883.

LINCOLN, P., WELCH, G., NASHEL, A., ILIE, A., AND FUCHS, H. 2009. Animatronic shader lamps avatars. *8th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2009*, 27–33.

LOPER, M. M., KOENIG, N. P., CHERNOVA, S. H., JONES, C. V., AND JENKINS, O. C. 2009. Mobile human-robot teaming with environmental tolerance. *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, 157–164.

LUBER, M., SPINELLO, L., AND ARRAS, K. O. 2011. People tracking in rgb-d data with on-line boosted target models. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3844–3849.

MONTEMERLO, M., PINEAU, J., ROY, N., THRUN, S., AND VERMA, V. 2002. Experiences with a mobile robotic guide for the elderly. *Proceedings of the National Conference on Artificial Intelligence*, 587–592.

OHYA, A., AND MUNEKATA, T. 2002. Intelligent escort robot moving together with human-interaction in accompanying behavior. *Proceedings of FIRA Robot World Congress*, 31–35.

PACCHIEROTTI, E., CHRISTENSEN, H. I., AND JENSFELT, P. 2006. Design of an office-guide robot for social interaction studies. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4965–4970.

PANG, W. C., BURHAN, AND SEET, G. 2012. Design considerations of a robotic head for telepresence applications. *International Conference on Intelligent Robotics and Applications (ICIRA)*, 131–140.

PRASSLER, E., BANK, D., KLUGE, B., AND HAGELE, M. 2002. Key technologies in robot assistants: Motion coordination between a human and a mobile robot. *Transactions on Control, Automation and Systems Engineering 4*, 56–61.

SEET, G., PANG, W. C., BURHAN, CHEN, I.-M., VIATCHESLAV, I., WILLIAM, G., AND WONG, C. Y. 2012. A design for a mobile robotic avatar - modular framework. *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*.

SEET, G. G. L., PANG, W. C., AND BURHAN. 2012. Towards the realization of maven - mobile robotic avatar. *The 25th International Conference on Computer Animation and Social Agents*.

TOPP, E. A., AND CHRISTENSEN, H. I. 2005. Tracking for following and passing persons. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2321–2327.

VENOLIA, G., TANG, J., CERVANTES, R., BLY, S., ROBERTSON, G., LEE, B., AND INKPEN, K. 2010. Embodied social proxy: mediating interpersonal connection in hub-and-satellite teams. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1049–1058.

WILLOW GARAGE, 2011. Texai remote presence system. Available online at `http://www.willowgarage.com/pages/texai/overview`.

YOSHIMI, T., NISHIYAMA, M., SONOURA, T., NAKAMOTO, H., TOKURA, S., SATO, H., OZAKI, F., MATSUHIRA, N., AND MIZOGUCHI, H. 2006. Development of a person following robot with vision based target detection. *32nd Annual Conference of IEEE Industrial Electronics*, 5286–5291.