Fusion of Automation and Teleoperation for Person-Following with Mobile Robots

Hemin Omer Latif, Student Member, IEEE, Nasser Sherkat, Ahmad Lotfi, Senior Member, IEEE

Abstract—Person-Following has been the focus of significant amount of research in the field of robotics recently. It is believed that detecting, recognizing and following people is one of the required functionalities for most future robots who will be sharing environments with their human companions. Fully automating these functionalities is considered as a significant challenge in developing any Person-Following system. This leads research to focus mostly on this challenge and divert from other challenges that coexist. A natural Person-Following functionality consists of a number of tasks that are required to be implemented in the system. However, in more realistic life scenarios, not all the tasks required for Person-Following need to be automated. Instead, some of these tasks can be operated by a human operator. In this paper, a novel taxonomy for Person-Following is introduced. The taxonomy presents all the tasks that are believed to exist in any Person-Following system. It also presents a number of likely cycles of tasks based on real life scenarios. Then, in order to provide a natural human-robot interaction means, fusion of automation and teleoperation for Person-Following is achieved by using TeleGaze. TeleGaze was previously developed by the authors as a means of natural human-robot interaction for teleoperation through eye gaze. The use of TeleGaze shows how all the tasks in Person-Following can be addressed and how to achieve a realistic fusion of automation and teleoperation.

Index Terms—Eye Tracking, Human-Robot Interaction, Mobile Robot Navigation, Person-Following, Teleoperation, TeleGaze.

I. INTRODUCTION

Person-Following (PF) is believed to be one of the main functionalities required for future robots to achieve natural Human-Robot Interaction (HRI). This functionality needs to be implemented so that future robots are able to interact with people in a variety of different environments under varying working and interacting conditions [1]. Therefore, PF is becoming an increasingly popular research topic in the field of robotics with, as reported in the literature, significant progress towards robust and reliable implementation of this functionality [2].

In general, the main challenges in any PF application is believed to be the task of tracking the Person-Of-Interest (POI) and the task of issuing the required motion commands to keep that POI within a desired distance of the robot. The first task is mostly implemented by using object tracking algorithms, with or without, some modifications [3] and the second task is implemented through different robot navigation algorithms, where certain functionalities such as obstacle avoidance can also be added, for example [4]. The aim of most related research though is the automation of these tasks in PF systems.

However, a complete PF system consists of a number of other tasks that raise a number of different challenges in different contexts of application. A natural PF system cannot be achieved without implementing all the tasks involved in the system. This, however, does not mean that all the tasks in PF need to be automated in order to achieve natural HRI. In fact, some of the tasks look more natural when they are not automated, but rather are operated by a human operator.

One of the natural modes of human-human interaction is through eye contacts and eye communications [5]. With the continues advancements in eye tracking technology, inputs from human eyes have been used in developing many Human-Computer Interaction applications [6] and therefore HRI applications too [7]. Eye tracking data is believed to be natural representation of human intentions and reactions [8]. Therefore, they are used widely in developing natural HRI applications with the aid of using intelligent user interfaces [9].

In this paper, in order to address most of the tasks required to be implemented in any PF system a special taxonomy is proposed. The PF taxonomy presents the list of tasks and the likely order of their implementation in any PF system. Also a number of likely cycles of task implementations are presented. In order to achieve a rather complete PF system in the form of natural HRI, inputs from human eyes are used to interact with a robotic agent. Teleoperation though eye gaze (TeleGaze) is integrated to an automated PF system implementing most of the tasks from the taxonomy of PF mentioned earlier in a natural form of HRI. The fusion of automation and teleoperation provides a realistic implementation of PF functionalities for mobile robots.

The paper is organized as follows: section 2 introduces the PF taxonomy with examples of likely scenarios. Section 3 covers the implementation of the taxonomy and how the tasks can be implemented using either automation or teleoperation. Then, section 4 includes a brief background on TeleGaze for
the sake of information integrity. In section 5, the integration of TeleGaze into PF is presented. Section 6 covers algorithms and apparatus used in developing the PF system and conclusions are in section 7.

II. THE TAXONOMY OF PERSON-FOLLOWING

The main challenge in any PF application is believed to be the challenge of keeping track of the person of interest (POI). This is, in most cases, addressed by developing, or modifying, an object tracking algorithm used to keep track of the POI [4]. Or, in some cases, to cope with variations in the interactions conditions fusion of cues and algorithms is used to address this challenge [10]. However, in order to build a rather complete PF system, it is required for research in the field to address a number of other challenging tasks too. Regardless of the complexity of the applications and the likely scenarios, a complete PF system consists of a number of tasks that each might raise a number of challenges during the course of interaction. Therefore, they need to be addressed in the development of any PF system.

In order to highlight the tasks involved in developing any PF system, the PF taxonomy is introduced. The taxonomy presents all the tasks that are required to be implemented in a natural form of HRI in any PF application. Also the taxonomy presents a number of likely interaction scenarios in the form of loops of interactions. Each loop of interaction consists of a number of tasks and the likely cycles of task implementation. The PF taxonomy is illustrated in Fig.1

![Fig. 1: The Person-Following (PF) Taxonomy.](image)

The terms tracking and following are used in the literature to refer to the same meaning and to different meanings interchangeably [1, 2, and 11]. Therefore, in order to standardize the use and the meaning of these two terms in PF applications, the term tracking in the taxonomy refers to the set of actions taking place in order to keep the POI in the vicinity of the robot without physically moving the robot. Whereas, the term following refers to the set of actions taking place in order to keep the POI in the vicinity of the robot by moving the robot. The later task requires distance information to keep the robot at a desired distance of the moving target which is the POI. Therefore, notice the difference between task two (start person-tracking) and task three (start person-following) in the taxonomy. Also notice the difference between Person-Following (PF) as the entire system and person-following (pf) as an individual task in the overall system.

Despite the ordinal representation of the tasks in the taxonomy, different loops in the taxonomy represent different interaction scenarios that are likely to happen in any PF application. Although, for instance, it is most likely that task two will be implemented once task one is implemented, task seven might be implemented after task one if a wrong person is registered. Therefore, the interaction loop that consists only of task one and task seven is a likely interaction scenario in real life PF applications. The mentioned scenario explains the
importance of the tasks in the PF taxonomy and how a PF application needs to address more than just the problem of tracking and following the POI.

In order to provide better understanding of likely interaction scenarios in PF applications, following is two examples of scenarios that involve other tasks in addition to the tasks of tracking and following the POI.

A. Scenario One:

An example of an interaction scenario is that a person gets registered in the PF system as the POI (task one) and the system starts tracking that person (task two). Then either due to change in interest, due to realizing that a wrong person is registered, or due to the reason that the system has lost the POI the system stops tracking that person (task six) and the person gets deregistered from the system as the POI (task seven). Therefore, for the system to perform PF, the cycle of tasks returns back to the task of registering a person (task one) and then continue from there to implement any other likely interaction loop based on the conditions of interaction.

As it can be seen from the previous interaction scenario, not all tasks in the PF taxonomy were involved in the scenario. Instead, a realistic interaction scenario, such as this one, could take place without invoking the task of following a person. Furthermore, even within the loop of this interaction scenario there are other possible scenarios that might take place as partial cycles of task implementations. The interaction loops of the tasks implemented in this scenario is illustrated in Fig. 2.

B. Scenario Two:

A person gets registered in the system as the POI (task one) and the system starts tracking (task two) and following (tasks three and four) the POI. Then the system stops following the POI (task five) but it still keeps tracking the person (task two). Or, it stops tracking the person (task six) but it keeps the registered person as the POI. In the former case, when the system stops following but keeps tracking the POI, the systems waits for restarting the person-following (task three). In the latter case however, the system needs to restart tracking the POI (task two). In both cases, person registration (task one) is not required as the same person is still registered in the system as the POI. The interaction loops of the tasks implemented in this scenario are illustrated in Fig. 3.

Fig. 3: The interaction loop and the cycle of tasks for the interaction scenario two. The same form of layout is used from Fig.1 for the sake of comparison.

Scenarios are believed to be very essential in designing any interactive system as they present stories about interactions [12]. The previous two interaction examples show how a number of tasks in a number of different likely cycles of implementation might be involved in a PF scenario. The presentation of the scenarios shows that how care must be taken not to limit the problem span of PF to the tasks of tracking and following only. Each of the tasks presented in the taxonomy requires attention in order to achieve natural HRI applications when implementing interaction applications such as the PF application.

III. THE IMPLEMENTATION OF THE PF TAXONOMY:

The combination of autonomous and non-autonomous functionalities in one application is a common approach in developing many HRI applications [13]. Some of the tasks in the taxonomy can be either operated by a human operator or automated to be implemented which means that not all those tasks require automation. In fact, some of them make more sense when they are operated by a human operator and not automated. One of the tasks for example that is most likely to require operation and not automation is registering the POI (task one). However, this does not mean that operating the task should be achieved in an artificial way and not considered from a natural HRI point of view. Implementing this task has been achieved in a number of different ways as reported in the literature so far such as a mouse selection, people detection [14], motion detection [11], or even a preregistered template such as a predetermined color [4]. This task however, when operated, needs to be implemented in a more natural way of HRI interaction [15].

Similar to the person registration task, starting person-tracking (task two), starting person-following (task three), stopping person-following (task five), stopping person-
tracking (task six) and finally person deregistration (task seven) can be operated in a PF application and not automated. Although some of these tasks are merged into one task in some applications such as starting person-tracking (task two) once the person registered (task one) and then starting person-following (task three) once person-tracking (task two) started. However, in a more realistic application each one of these tasks needs to be implemented once the conditions for their implementation are met and not as a group of tasks altogether. Therefore, an ideal PF application needs to deal with invoking each task separately from the other tasks in the taxonomy while it enables a natural HRI form of invoking each task. TeleGaze, which is introduced in the next section, is used as a natural means of HRI in developing and designing a rather realistic PF application.

IV. TELEGAZE FOR NATURAL HUMAN-ROBOT INTERACTION

In the effort of developing natural and intuitive means of HRI, the authors previously developed TeleGaze as a means of teleoperation through eye gaze. TeleGaze uses inputs from human eyes to enable a human operator to navigate a mobile robot from a remote location using an Intelligent User Interface. The TeleGaze interface enables both monitoring as well as controlling. Monitoring is achieved using real time images from a video camera mounted on the mobile robot. Controlling is achieved using inputs from the human operator’s eye to issue motion commands. Both monitoring and controlling are achieved with out any involvements of the operator’s hands as TeleGaze is essentially developed to reduce the amount of body engagement in teleoperation applications for mobile robots. If a human operator is able to navigate a mobile robot only using inputs from his/her eyes, then the hands of the operator are free from the navigation task, either partially or fully.

The TeleGaze interface is a powerful presentation of two layers of information on top of each other. The background layer is the real time images that come back from a video camera mounted onboard of the robot. This layer works as the feedback layer of the robotic platform and the status of the system. The background layer is augmented with a transparent layer in the foreground that enables controlling the robotic platform. The controlling layer is composed of a number of regions which are called action regions. The action regions are transparent regions each associated with a certain action command. Through the action regions, the operator is enabled to issue action commands required to move the robot, control the pan/tilt unit of the camera and control the TeleGaze interface itself. In order to issue a command, the operator needs to look at the action region associated with that particular command for a dwell time period of a third of a second. This is the time it approximately takes two consequent fixations to happen in the same action region.

An illustration of the layout of the TeleGaze interface with captions for each action region is shown in Fig. 4. To grab a better image of the two layers of the interface, an actual snapshot of the interface while in work is shown in Fig. 5.

Controlling the interface includes changing between different modes of interaction and different modes of operation. The two different modes of interaction are the interaction mode and the inspection mode. The interaction mode enables the operator to interact with the robot by issuing motion commands through the use of the action regions. The inspection mode enables the operator to use the interface to inspect the scene without issuing any commands except commands required to switch back to the interaction mode. The two different modes of operation are the TeleGaze mode and the PF mode. The TeleGaze mode enables the operator to interact with the robot using inputs from the eyes. The PF mode enables the operator to operate the robot in a PF mode. Once switched to the PF Mode the operator is enabled to switch back to the TeleGaze Mode using inputs from his/her eyes. Fig. 6 shows the TeleGaze interface in the PF Mode where only one action region is available to interact with and to switch back to the TeleGaze Mode if desired.
The experimental platform used in developing TeleGaze consists of a Wi-Fi enabled mobile robot, an eye tracking system together with a PC located in the remote teleoperation station. For more information on TeleGaze and the TeleGaze interface, the reader is recommended to refer to the authors’ previous works on TeleGaze [16, 17, and 18].

V. THE INTEGRATION OF TELEGAZE TO PERSON-FOLLOWING

As it was mentioned earlier TeleGaze enables the operator to operate the mobile robot in two different modes of operation. The TeleGaze mode enables teleoperation through human eye gaze. In other words, the robotic agent reads the intentions of its human partner by tracking its partner’s eye movements and corresponds to these eye movements in the form of action commands. The PF mode, however, enables the operator to change from a teleoperated mode to an automated PF following mode. This mode, based on the principle of understanding the operator’s intentions through eye movement data, enables the operator to select the POI by gazing at him/her for a certain period of time. The period of time in this mode is twice the dwell time used in the TeleGaze mode in order for the system to obtain more confirmation regarding the POI. Gazing at a person in the scene of the robot implicitly indicates that the operator is interested in following that person. This is a natural and intuitive implementation of registering the POI (task one) in the PF system. Once the POI is registered in the system, the system informs the operator by drawing a box surrounding the POI in the scene. When this task is completed, then the system starts tracking and following this person (tasks two and three). The dependent functionality of the system based on the interaction and operation modes via the TeleGaze interface is believed to achieve one of the basic principles of natural HRI which is implicit changes in modes of interaction [19].

If at any time during the course of PF, the operator decided to gain back control over the robot then all he/she needs to do is gazing at the action region associated with the operation mode. This changes the operation mode back to the TeleGaze mode where the operator can control the robot. In other words, stop following and tracking the POI (tasks five and six) and deregistering the POI (task seven). However, during the course of PF, if the robot lost the POI for any reasons, it keeps looking for him/her for a period of time. If the POI was found, then it starts following him/her again (tasks two and three). If the robot failed to find the POI, then it switches back to the TeleGaze mode where the operator teleoperates the robot and the POI gets deregistered (task seven). However, during the course of PF if the POI is lost, the robot keeps the registration of the lost person as the POI unless the operator intervene and change back to the TeleGaze mode or select a different person to be the POI.

VI. ALGORITHMS AND APPARATUS

For tracking purposes a basic version of the Camshift tracking algorithm was modified and implemented in OpenCV [20]. The Camshift algorithm enables a color based object tracking in real-time once a color blob is selected from the scene. It is not one of the objectives of this research to develop an object tracking algorithm. However, for the purpose of the PF application some modifications were made to the Camshift algorithm. Considering the sensitivity of the Camshift algorithm to rapidly changing scenery such as fast movements of the POI from one side of the scene to another, the modifications included expanding the searching span for the POI when (s)he is lost. This expands the functionality of the algorithm to search the whole scene for the POI and hence more chances to find the person if (s)he still exists in any part of the scene. However, the color blob that represents the POI needs to meet a minimum threshold of 10% of the image’s dimensions in order to be considered found and available for tracking.

A rather interesting modification to the algorithm is calculating the distance from the object been tracked. Solely depending on the images from one single video camera, the distance between the POI and the robot is kept at the initialization threshold. Once the POI is selected from the scene, the algorithm calculates the initial size of the color blob that represents the person. Then it keeps the distance from the person that keeps the color blob at the same initial size in the scene. This means any decrease in the size of the color blob leads to moving the robot towards the person and vice versa. The distance kept between the robot and the POI is highly flexible and depending on the initial distance when the POI is registered to the robot. Therefore, the task of following the person (task four) is implemented with highly natural but simple implementations of vision algorithms.

As far as the apparatus is concerned, the TeleGaze system consists of the eye tracking sub-system at one end, the robotic and vision sub-system at the other end and the TeleGaze interface in the middle. The TeleGaze interface and the software behind it work as a meeting point for the data flow from both ends of the system. TeleGaze is a platform independent system which can be implemented on any robotic platform equipped with active vision systems and with any eye
tracking platforms providing the required connectivity is achieved. For more information on the apparatus and the hardware architecture of TeleGaze the reader is recommended to refer to the author’s previous works on TeleGaze [16, 17, and 18].

VII. CONCLUSIONS
From the work presented in this paper, it can be concluded that the problem space of PF is not limited to one tracking algorithm or robotic navigation. There are a number of other tasks that need to be addressed as much as these two. Therefore, this paper presented taxonomy of PF for mobile robots. The taxonomy shows a number of different tasks that are involved in developing any PF application. Furthermore, implementing these tasks need to be done in a natural and intuitive way in order to achieve natural HRI. The implementation cycles of the tasks in the taxonomy might depend on the interaction scenario. Not all the tasks presented in the taxonomy might be invoked in all PF applications. However, the PF system needs to be developed so that it is capable of dealing with different tasks in the taxonomy and in different interaction scenarios. To achieve this aim, TeleGaze is integrated to a PF application. TeleGaze enables natural HRI and enables a robotic agent to understand the intentions of its human partner. The integration of TeleGaze to the PF application presented also shows an intuitive fusion of teleoperation and automation for HRI applications in real life scenarios.

To aid in the PF taxonomy, a standardized use of both terms person following and person tracking is proposed and used in the taxonomy. The authors recommend the presented standardization to be used in all future research related to PF. Finally, a novel technique for keeping a distance between the POI and the robot in PF applications is used. Through rather simple calculations and based on images from a single video camera, the initial distance between the robot and the POI is kept throughout the course of interaction using the results of the vision-based object tracking algorithm.

Although the taxonomy of PF is built based on generalized interaction scenarios, TeleGaze uses one single mode of interaction. To further generalize the application domain of PF and to enable more natural HRI, multi-modal interaction modes might be necessary. In a multi-modal interaction application, each task might be invoked with different modes of interaction. Therefore, future works of the authors investigate PF systems that address all the tasks in the presented taxonomy using a multi-modal interaction approach.

REFERENCES