Article

The International Journal of Robotics Research
2019, Vol. 38(14) 1581–1618

DOI: 10.1177/0278364919881683
journals.sagepub.com/home/ijr
SAGE

Person-following by autonomous robots: A categorical overview

Md Jahidul Islam, Jungseok Hong and Junaed Sattar

Abstract

A wide range of human–robot collaborative applications in diverse domains, such as manufacturing, health care, the entertainment industry, and social interactions, require an autonomous robot to follow its human companion. Different working environments and applications pose diverse challenges by adding constraints on the choice of sensors, degree of autonomy, and dynamics of a person-following robot. Researchers have addressed these challenges in many ways and contributed to the development of a large body of literature. This paper provides a comprehensive overview of the literature by categorizing different aspects of person-following by autonomous robots. Also, the corresponding operational challenges are identified based on various design choices for ground, underwater, and aerial scenarios. In addition, state-of-the-art methods for perception, planning, control, and interaction are elaborately discussed and their applicability in varied operational scenarios is presented. Then some of the prominent methods are qualitatively compared, corresponding practicalities are illustrated, and their feasibility is analyzed for various use cases. Furthermore, several prospective application areas are identified, and open problems are highlighted for future research.

Keywords

Person-following robot, human–robot interaction, human detection and tracking

1. Introduction

Person-following scenarios arise when a human being and an autonomous robot collaborate on a common task that requires the robot to follow the human. Usually, the human leads the task and cooperates with the robot during task execution. An example application would be the service robots that are widely used in industrial applications, e.g., in manufacturing, warehousing, and health care. The use of companion robots in surveillance, social interaction, and medical applications has also flourished over the last decade. Numerous new applications are also emerging in the entertainment industry as robots are becoming more accessible for personal use.

Based on the operating medium, person-following can be categorized into ground, underwater, and aerial scenarios. A ground service robot following a human while performing a cooperative task is the canonical example of person-following. Such assistant robots are being used in many domestic and industrial applications (Piaggio-Fast-Forward, 2017) and in health care (Ilias et al., 2014; Iribe et al., 2011). Moreover, diver-following robots are useful in submarine pipeline and shipwreck inspection, marine life and seabed monitoring, and many other underwater exploration activities (Islam et al., 2018a; Miskovic et al., 2015; Sattar and Dudek, 2009b). Furthermore, the use of person-following aerial robots (Mueller et al., 2016; Naseer et al., 2013; Pestana et al., 2014) has flourished over the last decade in the entertainment industry (Skydio, 2018) as quadcopters have become quite popular for filming outdoor activities, such as mountain climbing, biking, surfing, and many other sporting endeavors.

In all these applications, following a person is not the primary objective of the robot, yet it is vital for achieving the overall success of the primary operation. Robust techniques to enable person-following are thus of significant importance in the repertoire of robotic behaviors. The major computational components of a person-following system are perception, planning, control, and interaction. The design of each of these components largely depends on the choice of sensors and the degree of autonomy required for the robot. Additionally, different scenarios

Interactive Robotics and Vision Laboratory, Department of Computer Science and Engineering, University of Minnesota, USA

Corresponding author:
Md Jahidul Islam, Interactive Robotics and Vision Laboratory, Department of Computer Science and Engineering, University of Minnesota, Twin Cities, MN, USA.
Email: islam034@umn.edu

(i.e., ground, underwater, and aerial) pose different operational challenges and add constraints on the construction and dynamics of the robot. These factors are determined by specific application requirements, which make it difficult to design a generic person-following methodology.

Attempts to develop person-following robots for a wide range of applications have resulted in a variety of different methodologies. In particular, computer vision and robotics researchers have been developing person-following methodologies for ground robots since the nineties (Azarbayejani and Pentland, 1996; Darrell et al., 1998; Wren et al., 1997). Initially seen as a special case of object tracking, person-following by autonomous robots soon became a challenging problem of its own as many industrial applications started to flourish (Balan et al., 2005; Cu et al., 2013; Ess et al., 2008; Pairo et al., 2013). Recently, other aspects of the problem, such as human–robot interaction, social awareness, and the degree of autonomy, are also receiving attention from researchers (Dewantara and Miura, 2016; Triebel et al., 2016). The advent of underwater and aerial applications has added other dimensions to this growing field (Mueller et al., 2016; Naseer et al., 2013; Sattar and Dudek, 2009b). Different mediums and a diverse set of operational considerations often demand application-specific design for a person-following robot. However, certain design issues, underlying algorithms, and methods of human–robot interaction, among others, remain mostly generic for all person-following scenarios. An elaborate discussion of these aspects, with a comparison of different approaches and the state-of-the-art techniques would greatly help current and future researchers.

This paper outlines various aspects of the person-following problem and provides a comprehensive overview of the existing literature. In addition, different issues pertaining to robot and algorithmic design are identified, operational scenarios are illustrated, and qualitative analyses of the state-of-the-art approaches are presented. Specifically, the contributions of this paper are the following:

- A categorization of the person-following problem is presented based on various attributes, such as the medium of operation, choice of sensors, mode of interaction, and degree of autonomy. Operational scenarios for each category are then discussed, along with the relevant applications.
- Additionally, for different person-following scenarios, key design issues are identified, the underlying assumptions are discussed, and state-of-the-art approaches to cope with the operational challenges are presented.
- Subsequently, an elaborate discussion of the underlying algorithms of different state-of-the-art approaches for perception, planning, control, and interaction are presented. The attributes and overall feasibility of these algorithms are qualitatively analyzed and then compared based on various operational considerations.
- Furthermore, several open problems for future research are highlighted, along with their current status in the literature.

2. Categorization of autonomous person-following behaviors

Person-following behaviors by autonomous robots can be diverse depending on several application-specific factors, such as the medium of operation, choice of sensors, mode of interaction, granularity, and degree of autonomy. The design and overall operation of a person-following robot mostly depend on the operating medium, e.g., ground, underwater, and aerial. Other application-specific constraints influence the choice of sensors, mode of interaction (explicit or implicit), granularity, and degree of autonomy (full or partial). In this paper, *explicit* and *implicit* interactions refer to direct and indirect human–robot communication, respectively. In addition, the term *granularity* is defined as the number of humans and robots involved in a person-following scenario.

Based on these attributes, a simplified categorization of autonomous person-following behaviors is depicted in Figure 1. The rest of this paper is organized by following the categorization based on the medium of operation, while other attributes are discussed as subcategories.

2.1. Ground scenario

Domestic assistant robots (Piaggio-Fast-Forward, 2017) and shopping-cart robots (Nishimura et al., 2007) are the most common examples of person-following unmanned ground vehicles (UGVs). Their usage in several other industrial applications (The-5elementsrobotics, 2014), and in health care has also increased in recent times (Ilias et al., 2014; Iribe et al., 2011; Tasaki et al., 2015). Figure 2 illustrates typical person-following scenarios for ground robots. The UGV uses its camera and other sensors to detect the person in its field of view. Once the position and distance of the person are approximated, path planning is performed, and the corresponding control signals are generated in order to follow the person. Details of these operations and the related state-of-the-art methodologies will be discussed later in this paper. The following discussion expounds various design issues and related operational constraints based on the choice of sensors, mode of interaction, granularity, and degree of autonomy.

2.1.1. Choice of sensors. Most person-following UGVs are equipped with cameras and the perception is performed through visual sensing. Other sensors are used to accurately measure the distance and activities (walking, waving hands, etc.) of the person for safe navigation and interaction. The choice of sensors is often determined by the operating environment, i.e., indoors or outdoors. For example, RGBD sensors are very effective in indoor environments; in

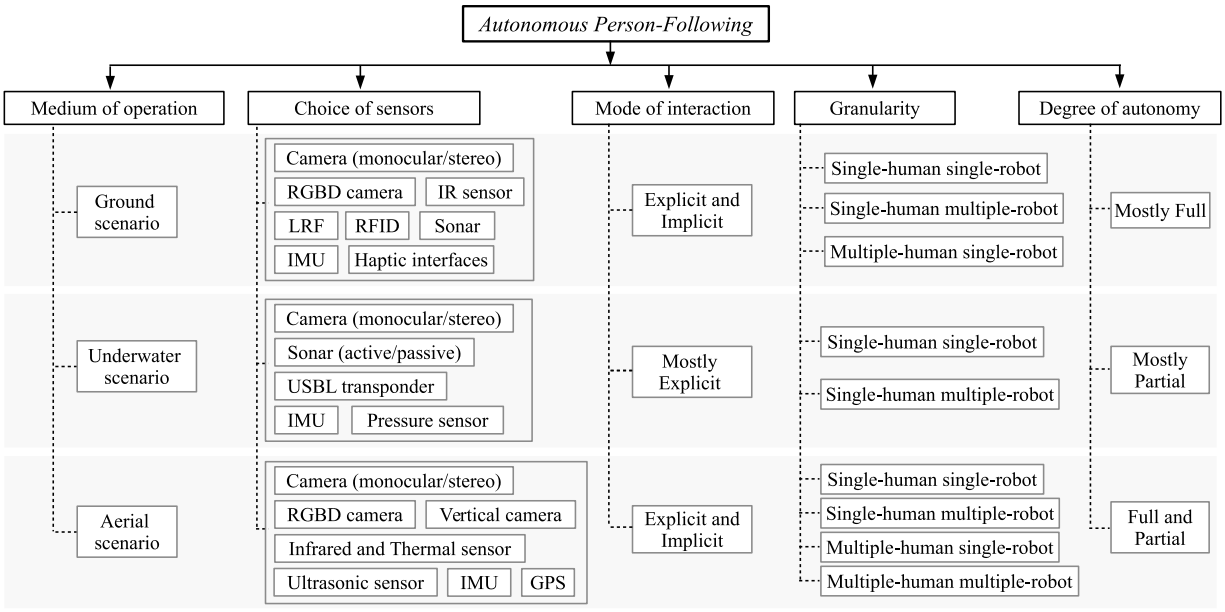


Fig. 1. A categorization of autonomous person-following behaviors based on five major attributes: medium of operation, choice of sensors, mode of interaction, granularity, and degree of autonomy.
GPS: Global Positioning System; IMU: inertial measurement unit; IR: infrared; LRF: laser range finder; RFID: radio-frequency identification; RGBD: RGB-depth; USBL: ultrashort baseline.

addition to having a regular RGB camera, they are equipped with an infrared sensor to provide the associated depth information (Mi et al., 2016). Therefore, both the position and distance of the person can be measured with good accuracy. However, since infrared sensors perform poorly in the presence of sunlight, they are not good choices for outdoor environments. The use of stereo cameras can get rid of this problem, as depth information can be approximated by using triangulation techniques (Chen et al., 2017b; Satake et al., 2013). Laser range finders (LRFs) are also widely used by person-following UGVs (Chung et al., 2012; Susperregi et al., 2013). These sensors provide a cluster of directional distance measures, from which the presence and distance of a person can be approximated. Moreover, some applications use ultrawide band (UWB) (Laneurit et al., 2016), radio-frequency identification (RFID) tags (Germa et al., 2010; Kulykukin et al., 2004), and haptic interfaces (Ghosh et al., 2014) as exteroceptive sensors.

Proprioceptive sensors, such as inertial measurement units (IMUs), are used to keep track of the robot's relative motion and orientation information (Brookshire, 2010) for smooth and accurate navigation. Researchers have also explored the use of wearable IMUs for modeling the human walking gait (Cifuentes et al., 2014); this is useful for differentiating humans from other moving objects.

Person-following UGVs typically use a number of sensors in order to ensure robustness and efficiency. Standard sensor fusion techniques are then adopted to reduce the uncertainty in sensing and estimation (Cifuentes et al., 2014; Luo et al., 2009; Susperregi et al., 2013). A summary of the key design issues based on different choices of sensors for person-following UGVs is presented in Table 1.

(a) Indoors. **(b)** Outdoors.

Fig. 2. Typical person-following scenarios for ground robots: (a) TurtleBot (Willow-Garage, 2011) following a person in an indoor setting; (b) Gita cargo robot (Piaggio-Fast-Forward, 2017) following a person outdoors.

2.1.2. Mode of interaction. It is ideal for a person-following robot to interact with the human user in a natural way. Even if the robot is not designed to interact with the person directly, implicit (i.e., indirect) interactions exist, since the person is aware that the robot is following (Hu et al., 2014). This awareness is important for the overall success of the operation because the person can simplify the robot's task in many ways, e.g., by making smooth turns, avoiding obstacles, and walking at a reasonable speed. Additionally, the robot needs to maintain a safe

Table 1. Choice of sensors and corresponding design issues for person-following unmanned ground vehicles.

Sensor	Data	Challenges or limitations	Usage or operation	Selected references
Monocular camera	RGB image	Low visibility; lighting variation	Computer vision-based algorithms are used for detection and tracking	Guevara et al. (2016); Isobe et al. (2014); Kobilarov et al. (2006); Kwon et al. (2005); Ma et al. (2008); Pierre (2018)
Stereo camera	RGB image	Low visibility; lighting variation	In addition to RGB image-based detection and tracking, stereo triangulation techniques are used to approximate the associated depth information	Brookshire (2010); Chen et al. (2017b); Hu et al. (2014); Itoh et al. (2006); Luo et al. (2009); Satake and Miura (2009); Satake et al. (2012, 2013); Takemura et al. (2007)
RGBD camera	RGBD data	Presence of sunlight	In addition to RGB image-based detection and tracking, distance of the person is approximated using the depth data	Basso et al. (2013); Cosgun et al. (2013); Doisy et al. (2012); Masuzawa et al. (2017); Mi et al. (2016); Munaro et al. (2013); Wang et al. (2017)
Laser range finder	Planner distance measures	Presence of transparent (e.g., glass) or dark surfaces	Person's body is detected from a cluster of distance measures	Alvarez-Santos et al. (2012); Cai and Matsumaru (2014); Cosgun et al. (2013); Jung and Yi (2012); Leigh et al. (2015); Pairo et al. (2013); Shaker et al. (2008)
Sonar	Directional distance measures	Specular reflections; crosstalk	Presence and distance of a person is detected from the directional distance measures	Gascueña and Fernández-Caballero (2011); Itoh et al. (2006); Peng et al. (2016)
Radio-frequency identification	RF signal	Presence of interfering signals; limited range; ensuring privacy	Person carrying an RFID tag is tracked by following the direction of the RFID signal source	Germa et al. (2010); Kulykukin et al. (2004)
Inertial measurement unit (IMU)	IMU data	Precision; drift	Robot's relative orientation, angular and linear velocity, and acceleration are estimated for motion control	Brookshire (2010); Cifuentes et al. (2014)

Table 2. Challenges and responsibilities involved in implicit and explicit human–robot interactions for person-following UGVs.

Interaction	Challenges for the robot	Responsibilities of the person	Selected references
Implicit	Maintaining safe distance and speed; socially aware spatial and motion conduct planning	Walking at reasonable speed; avoiding obstacles; taking smooth turns	Ferrer et al. (2013); Gockley et al. (2007); Honig et al. (2018); Kulykukin et al. (2004); Kuno et al. (2007); Liem et al. (2008); Matsumaru et al. (2005); Triebel et al. (2016)
Explicit	*Additional challenges*: recognizing and decoding instruction commands; planning and acting based on given instructions	*Additional responsibility*: communicating clearly based on the predefined scheme	Cosgun et al. (2013); Doisy et al. (2013); Ghosh et al. (2014); Hirai and Mizoguchi (2003); Marge et al. (2011); Triebel et al. (2016)

distance and plan a socially aware motion trajectory while following the person (Granata and Bidaud, 2012; Triebel et al., 2016). A number of user studies have explored different aspects of implicit interactions, including but not limited to the robot's spatial conduct (Fleishman et al., 2018), preferred following angles (Shanee et al., 2016), turning behaviors (Hu et al., 2014), and socially aware motion conduct (Honig et al., 2018; Triebel et al., 2016). An elaborate discussion of these aspects is provided later in this paper (see Section 3.3.2).

Explicit interactions refer to direct human–robot communication. In many practical applications, a human instructs the UGV to perform certain tasks, such as changing its motion or speed, taking photographs, or making phone calls. These instructions are typically communicated using voice commands (Fritsch et al., 2004), hand gestures (Doisy et al., 2013; Marge et al., 2011), or haptic interfaces (Ghosh et al., 2014; Park and Howard, 2010). Moreover, some smart carts and autonomous luggage robots allow users to interact using smartphone applications (DiGiacomcantonio and Gebreyes, 2014). Explicit human–robot interaction is essential for most person-following ground applications; however, additional computational capabilities are required in order for the UGVs to understand human instructions and spontaneously interact in a natural manner. Table 2 highlights the challenges and responsibilities involved in different forms of human–robot interactions for person-following UGVs.

2.1.3. Granularity. Most domestic applications involve a single robot accompanying a single person. Interacting with a specific person is also common, particularly for accompanying older people and people with disabilities (Ilias et al., 2014; Kulykukin et al., 2004; Liem et al., 2008). The most important features of these robots are the social and interactive skills needed to make the accompanying person feel safe and attended to. In industrial applications, however, robustness and performance are more important, relative to social aspects (Cosgun et al., 2013; Germa et al., 2010). These robots typically assist a single person in a dynamic multi-agent environment, i.e., with the presence of other humans and robots.

A robot can also accompany a group of people by following the *center of attention* of the group (Basso et al., 2013; Chen et al., 2017b). However, this can be challenging if people move in random directions. An anchor person is generally specified to the robot beforehand to interact with the robot and help it to navigate. In such cases, the robot uses features specific to the anchor person for tracking while interacting with the group as a whole. Since interacting with a group of people can be challenging, service robots are often equipped with user interfaces for easy and effective human–robot interaction. Furthermore, a number of independent robots can assist a single person in a common task, given that the person is responsible for synchronizing their activities. Although swarm-like multi-robot

Table 3. Choice of sensors and corresponding design issues for person-following underwater robots.

Sensor	Data	Challenges or limitations	Usage or operation	Selected references
Monocular camera	RGB image	Poor lighting conditions and visibility; suspended particles; color distortions	Computer vision-based algorithms are used for detection and tracking	Islam and Sattar (2017); Islam et al. (2018a); Sattar and Dudek (2006)
Stereo camera	RGB image	Poor lighting conditions and visibility; suspended particles; color distortions	In addition to RGB image space based detection and tracking, stereo triangulation techniques are used to approximate the associated depth	Corke et al. (2007); Stilinović et al. (2015)
Active sonar	Directional distance measures	Noisy reading; scattering and reverberation	Diver's presence and distance are approximated from the directional distance measures	DeMarco et al. (2013); Mandic et al. (2016)
Passive sonar	Frequency responses	Extremely noisy reading; limited coverage	Frequency responses of the sound wave generated by the diver are used for detection	Gemba et al. (2014); Hari et al. (2015)
Ultrashort baseline (USBL) transponder	Acoustic pulse	Presence of a paired USBL transceiver; noisy reading	Diver's position information is estimated by communicating with the transceiver	Corke et al. (2007); Mandic et al. (2016); Miskovic et al. (2015)
Inertial measurement unit (IMU)	IMU data	Precision; drift	Robot's relative orientation, angular and linear velocity, and acceleration are estimated for motion control	Miskovic et al. (2015)
Pressure sensor	Depth measurement	Sensitiveness to temperature	Depth of the robot is approximated using the measured external pressure	Corke et al. (2007)

Table 4. Choice of sensors and corresponding design issues for person-following unmanned aerial vehicles.

Sensor	Data	Challenges or limitations	Usage or operation	Selected references
Front camera	RGB image	Low visibility; lighting variation	Computer vision algorithms are used for detection and tracking	De Smedt et al. (2015); Gaszczak et al. (2011); Pestana et al. (2014); Skydio (2018)
Vertical camera	RGB image	Lack of ground textures	Ground textures and visible features are used for stabilization	Lugo and Zell (2014)
RGB-depth camera	RGB-depth data	Presence of sunlight	In addition to RGB image space based detection and tracking, distance of the person is approximated from depth data	Lichtenstern et al. (2012); Naseer et al. (2013)
Ultrasonic sensor	Distance measure	Limited range	Vertical displacement is measured from the distance measures	Barták and Vykovský (2015); Lugo and Zell (2014)
Infrared and thermal sensor	Thermal image	Low resolution	Thermal radiation of a person is detected	Kumar et al. (2011); Rudol and Doherty (2008)
Inertial measurement unit (IMU)	IMU data	Precision; drift	Flight controllers use the orientation, angular speed, acceleration, and magnetic field information for navigation	Barták and Vykovský (2015); Lugo and Zell (2014)
Global Positioning System	Global position, speed, and time	Signal strength; accuracy in indoor settings	The triangulated positioning information is used by the control loop for smart navigation	DJI (2015); Rudol and Doherty (2008)

cooperation or non-cooperative multi-agent synchronization (Chen et al., 2017e) is possible, these frameworks are rather resource-demanding and not commonly adopted in person-following applications.

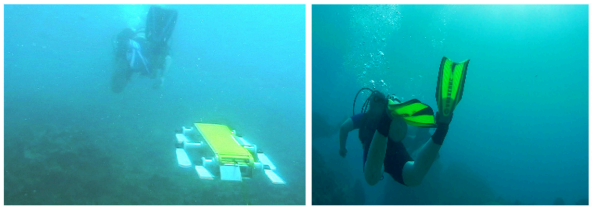
2.1.4. Degree of autonomy. A major advantage of using person-following robots is that it eliminates the need for dedicated teleoperation. Since autonomous ground navigation is relatively less challenging than underwater or aerial scenarios, person-following UGVs are typically designed to have fully autonomous behavior (Leigh et al., 2015). Some applications, however, allow partial autonomy for UGVs that perform very specific tasks, such as assisting a nurse in an operating room (Ilias et al., 2014) or serving food at a restaurant (Pieska et al., 2013). These service robots follow their companions around within a predefined operating area and provide assistance by carrying or organizing equipment, serving food, etc. While doing so, they may take human inputs for making navigation decisions, such as when to follow, on which side to stay, where or when to wait, which objects to carry or organize, etc.

Such semi-autonomous behaviors for UGVs are adopted in robot-guiding applications as well, e.g., guiding a visually impaired person (Ghosh et al., 2014) or tour guiding at a museum or shopping mall (Burgard et al., 1998; Kanda et al., 2009). Although robot-guiding is not strictly a person-following application, it shares a similar set of features and operational challenges for assisting a human companion. In particular, features such as socially aware planning, some aspects of explicit interaction, navigating through crowds while guiding or leading people, etc., are closely related to person-following applications. Readers are referred to Table 6 in Section 4 for an organized and annotated collection of the person-following (and relevant) literature.

2.2. Underwater scenario

Underwater missions are often conducted by a team of human divers and autonomous robots, who cooperatively perform a set of common tasks (Islam et al., 2018c; Sattar et al., 2008). The divers typically lead the tasks and interact with the robots, which follow the divers at certain stages of the mission (Islam et al., 2018a). These situations arise in important applications, such as the inspection of ship hulls and submarine pipelines, the study of marine species migration, and search-and-rescue or surveillance operations. In these applications, following and interacting with the companion diver (Islam et al., 2018c) is essential because fully autonomous navigation is challenging, owing to the lack of radio communication and global positioning information underwater. Additionally, the human-in-the-loop guidance reduces operational overhead by eliminating the necessity of teleoperation or complex mission planning a priori.

Figure 3 illustrates a scenario in which an autonomous underwater vehicle (AUV) is following a scuba diver



(a) Underwater robot following a diver. **(b)** Diver seen from the robot's camera.

Fig. 3. A typical diver-following scenario for an underwater robot during a reef exploration task.

during an underwater mission. The operational complexities and risks involved in underwater applications are generally much greater than those in ground applications (Sattar and Dudek, 2006). The following sections discuss these operational challenges and the related design issues based on the choice of sensors, mode of interaction, granularity, and degree of autonomy.

2.2.1. Choice of sensors. Underwater diver-following robots usually rely on vision for tracking, owing to the bandwidth limitations of acoustic modems. In addition, it is undesirable to be intrusive and disruptive to the ecosystem (Slabbekoorn et al., 2010). Cameras, being passive sensors (i.e., they do not emit energy), are thus preferred over active sensors. Additionally, the use of stereo cameras is effective in approximating the relative distance of a diver or other targets (Corke et al., 2007; Stilinović et al., 2015); standard computer vision-based techniques are then utilized for visual tracking. Although visibility can be a challenge, there is usually ample natural daylight at depths (typically 20–25 m) where human beings can dive and remain for extended periods of time without using specialized equipment.

However, visual data gets noisy, owing to challenging marine conditions (Sattar and Dudek, 2006) arising from such factors as color distortions, lighting variations, or suspended particles. Consequently, robust visual detection and tracking become extremely difficult. Passive sonars, such as hydrophones are useful in such scenarios (Gemba et al., 2014; Hari et al., 2015). Active sonars are also used for diver-following in unfavorable visual conditions (DeMarco et al., 2013; Mandic et al., 2016). They are particularly useful when a robot loses the diver from its field of view and tries to rediscover the diver; once rediscovered, the robot can switch back to visual tracking. Conversely, ultrashort baseline (USBL) is often used for global positioning of underwater robots and remotely operated vehicles (ROVs). A USBL transponder (mounted on the robot) communicates with a USBL transceiver (mounted on a pole under a ship or a boat) using acoustic signals. Phase-differencing methods are then used by the USBL to calculate positioning information (range, angle, etc.). The robot uses this information for navigation and tracking divers or other objects of interest.

Proprioceptive sensors, such as IMUs, are also used by underwater robots for internal state estimation (Miskovic et al., 2015); in addition, pressure sensors are used for measuring the depth of the robot (from the surface) using external pressure (Corke et al., 2007). The depth information is useful for the depth-control and altitude-control modules of the robot. Moreover, inertial navigation systems and other navigation sensors can be used to determine the robot's instantaneous pose and velocity information; however, these systems drift, thus requiring repeated correction using secondary sensing systems. Table 3 summarizes the challenges and operational issues based on different choices of sensors for person-following underwater robots.

2.2.2. Mode of interaction. Since truly autonomous underwater navigation is still an open challenge, explicit interaction with the accompanying diver is crucial for diver-following robots. In particular, in complex missions, such as surveillance and rescue operations, robots can dynamically adjust their mission parameters by regularly communicating with the diver. In addition, some underwater exploration and data collection processes require close human supervision. In these scenarios, the divers typically instruct the robot to perform certain tasks (record snapshots, take samples, etc.) in different situations (Islam et al., 2018c). Although such communication paradigms are fairly straightforward in terrestrial settings, these are rather complex undertakings for underwater robots.

A number of communication frameworks have been proposed for underwater human–robot interaction. In RoboChat (Dudek et al., 2007), divers use a set of "AR-Tag" markers to display a predefined sequence of symbolic patterns to the robot; these patterns are then mapped to a set of grammar rules defined for the language. A major limitation of such marker-based frameworks is that the markers need to be carried along and used in the correct order to produce instruction commands for the robot. Although the number of required markers can be reduced by incorporating additional shapes or motion signs with each marker (Sattar et al., 2007; Xu et al., 2008), this framework still involves a significant cognitive load on the diver. A simpler alternative is to use hand gestures to communicate with the robot (Chiarella et al., 2018; Islam et al., 2018b). This comes more naturally to divers because they already communicate with each other using hand gestures. Conversely, robots can communicate emergency messages (e.g., low battery) and periodic updates to the diver using an on-board screen, flashing lights, etc.

The social and behavioral aspects of underwater missions are limited (Wu et al., 2015). However, implicit diver–robot interactions are vital for ensuring the robot's safety and the overall success of the operation. The associated cognitive load on the divers is another important

consideration for designing an interaction framework (Chiarella et al., 2015; Islam et al., 2018b).

2.2.3. Granularity. As mentioned, the applications envisioned for underwater diver-following robots usually require a team of divers and robots. In most cases, each robot is assigned to one leader (usually a diver) who guides the robot during a mission (Islam et al., 2018a). The leader can be another robot as well. For instance, a robot can follow another robot, which is following a diver; such operations are referred to as robot convoying (Shkurti et al., 2017). Robot convoying is useful when there are more robots than divers. Additionally, it is often more convenient and safer than having a number of robots follow a single diver. Underwater robots are usually not assigned to follow more than one diver because this requires complex motion planning; also, interacting with a number of humans simultaneously can be computationally demanding and often problematic.

2.2.4. Degree of autonomy. Since underwater missions are strictly constrained by time and physical resources, most diver-following applications use semi-autonomous robots that take human inputs to make navigation decisions when needed. This reduces the overhead associated with underwater robot deployment and simplifies associated mission planning. For simple applications, diver-following robots are typically programmed to perform only some basic tasks autonomously, e.g., following the diver, hovering, or taking snapshots. These programs and associated parameters are numbered and made known to the robot (and diver) beforehand. The diver leads the mission and instructs the robot to execute (one of) these programs during operation. For instance, the robot might be instructed to follow the diver to the operation zone, then to hover at a particular location of interest, take pictures, and eventually follow the diver back at the end of the mission. This interactive operational loop is very useful for simple applications, such as exploration and data collection (Islam et al., 2018c). However, more autonomous capabilities are required for complex applications, such as surveillance or monitoring the migration of marine species. ROVs are typically deployed for these critical applications; these are connected to a surface vehicle (usually a ship or a boat) via an umbilical link that houses communication cables and an energy source, to enable power and information transfer.

2.3. Aerial scenario

Unmanned aerial vehicles (UAVs) are traditionally used for surveillance, industrial, and military applications. More recently, UAVs have become more accessible and popular for entertainment purposes and in the film industry. They are very useful for capturing sports activities, such as climbing or skiing, from a whole new perspective (Higuchi et al., 2011; Skydio, 2018; Staaker, 2016) without the need for teleoperation or a full-scale manned aerial vehicle. Another interesting application is to use person-following UAVs to provide external visual imagery, which allows athletes to gain a better understanding of their motions (Higuchi et al., 2011). These popular use cases have influenced significant endeavors in research and development for affordable UAVs, and they have been at the forefront of person-following aerial drone industry in recent times.

Figure 4 illustrates a typical person-following scenario for a UAV. The operating time for UAVs is usually much shorter than for ground and underwater scenarios, e.g., less than half an hour to a few hours per episode, owing to limited battery life. The person launches the take-off command at the beginning of each episode and then commands the UAV to follow (and possibly to take snapshots) while he or she is performing some activities. The person makes the landing command after a reasonable amount of time, ending the episode. It is common to carry a number of portable batteries or quick chargers to capture longer events. The following sections discuss other operational considerations and related design issues based on the choice of sensors, mode of interaction, granularity, and degree of autonomy.

2.3.1. Choice of sensors. As the mentioned applications suggest, person-following UAVs are equipped with cameras for visual sensing. Usually, a front-facing camera is used for this purpose, while an additional low-resolution vertical camera (i.e., facing down) is used as an optical flow sensor. The vertical camera uses ground textures and visible features to determine the UAV's ground velocity and ensure stabilization. Owing to the constraints on cost, weight, size, and battery life, the use of other exteroceptive sensors is often limited to consumer-grade person-following UAVs. The Parrot ARDrone 2.0 (Parrot, 2012), for instance, only uses cameras (front and vertical) as exteroceptive sensors; these UAVs weigh less than a pound and cost approximately two hundred US dollars. Conversely, with a 4k resolution camera and a three-axis mechanical gimbal, the DJI Mavic drones (DJI, 2016) weigh 700–850 grams and cost approximately a thousand US dollars.

However, UAVs used in industrial, military, and other critical applications can accommodate several high-resolution cameras, range sensors, stereo cameras, etc. For instance, Inspire 2.0 (DJI, 2015) drones have additional upward-facing infrared sensors for upward obstacle avoidance, ultrasonic sensors, and camera gimbals for stable forward vision. While these drones weigh about 6–8 pounds and cost a few thousand US dollars, they offer the robustness and reliability required for critical applications. Moreover, infrared and thermal cameras (Kumar et al., 2011) are particularly useful in autonomous human surveillance and rescue operations in darkness and during adverse weather. These sensors provide low-resolution thermal images (Rudol and Doherty, 2008), which are used to localize moving targets (e.g., people) in darkness. While multiple high-resolution stabilized cameras are useful in these

applications, manufacturers of person-following UAVs tend to avoid using other exteroceptive sensors and try to balance the trade-off between cost and battery life. For instance, although laser scanners are widely used by UAVs for surveying tasks involving mapping and localization (Huh et al., 2013; Tomic et al., 2012), these are not commonly used for person-following applications.

Lastly, proprioceptive sensors are used mainly by flight controller modules. For instance, IMUs measure three-axis rotations and acceleration while an optical flow sensor measures the horizontal (ground) velocity of the UAV. Additionally, ultrasonic and pressure sensors measure altitude and vertical displacements of the UAV (Barták and Vykovský, 2015). Flight controller modules use these sensory measurements to estimate the UAV pose and eventually control its position and trajectory during flight. Hence, these sensors are critical for the overall successes of the operations. Additionally, advanced UAVs make use of Global Positioning System (GPS)receivers within the navigation and control loop, allowing for smart navigation features, such as maintaining a fixed position or altitude. Table 4 summarizes the usage and challenges of different sensors used by person-following UAVs.

2.3.2. Mode of interaction. Since the per-episode operating time for UAVs is significantly shorter than that of UGVs and AUVs, their take-offs and landings are frequent. This requires that the person be aware of the UAV's location and available battery at all times in order to facilitate smooth person-following and ease the landing processes. Additionally, for a UAV paired to a user application via a wireless local area network (WLAN), the person being followed should not venture outside the WLAN range. Furthermore, the person can positively influence the behavior of the UAV by understanding the underlying algorithms, e.g., by knowing how the UAV navigates around an obstacle, how rediscovery happens when the target person is lost, etc. While these positive influences via implicit interactions are important for person-following in general, they are more essential in the aerial scenario.

As mentioned earlier, implicit interaction incurs additional cognitive loads on the user. To this end, explicit interactions and commands can simplify the task of controlling the UAV. Most commercial UAVs can be controlled via smart devices (DJI, 2016; Skydio, 2018), proprietary controllers, wearable beacons, etc. Moreover, hand-gesture-based interaction is particularly popular in personal applications where the UAV flies close to the person at a low altitude (Naseer et al., 2013). Typical instructions involve changing the robot's position or initiating a particular task, such as to start circling around the person, start or stop video recording, or make an emergency landing. Nevertheless, hand-gesture-based interaction with UAVs can be quite challenging when the UAV flies at a high altitude; these challenges are elaborately discussed later in this paper (Section 3.3.1).

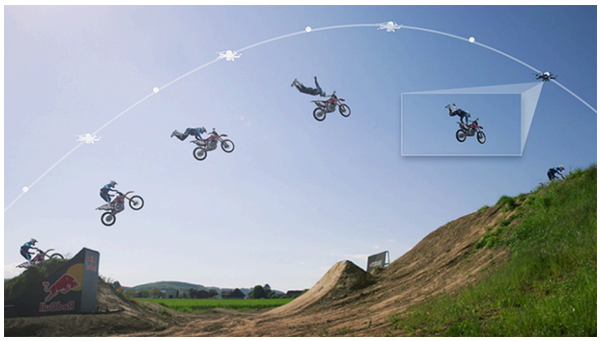


Fig. 4. Unmanned aerial vehicle filming a sport activity while intelligently following an athlete (Wellbots, 2015).

2.3.3. Granularity. As with the ground and underwater scenarios, a single UAV follows a single person in most commercial and personal applications. Owing to the increasing popularity of these applications, research studies have also concentrated largely on this interaction paradigm (Chakrabarty et al., 2016; Lugo and Zell, 2014; Pestana et al., 2014). However, a single UAV often cannot fulfill certain application requirements, such as capturing an event from different viewpoints or over a long period of time. Hence, critical applications, such as search-and-rescue operations, require several UAVs to follow a team (Cacace et al., 2016) and often share a cooperative task (e.g., covering a certain search perimeter). Moreover, a group of cooperative UGVs is more effective for crowd control (Minaeian et al., 2016), than is a single UAV.

While the integration of a number of person-following UAVs can overcome certain limitations of using a single UAV, controlling and interacting with a number of UAVs can become increasingly difficult. The cognitive load on the users is significantly increased as they need to worry about the battery life, take-off and landing, position, movement, etc., of each UAV. Although it is theoretically possible to interact with several UAVs separately and as a group using hand gestures or smart devices, it is not practical for most personal applications. For critical applications, however, UAVs with more advanced autonomous capabilities are used to reduce the cognitive load on the person. In fact, advanced UAVs have features that allow interactions with several persons, who share the cognitive load in complex operations. For instance, the camera gimbals of Inspire 2.0 (DJI, 2015) can be controlled independently (by a person) while it is interacting with a different person.

2.3.4. Degree of autonomy. Unlike ground scenarios, partial autonomy is preferred over full autonomy in most applications for person-following UAVs. The person usually uses a smartphone application for take-offs, positioning, and landing. Then, the UAV switches to autonomous mode and starts following the person. During operation, the person typically uses a smartphone application or hand gestures to communicate simple instructions for moving the

UAV around, taking snapshots, etc. If the UAV loses visual on the person, it hovers until rediscovery is made. UAVs are also capable of emergency landing by themselves if necessary (e.g., when the battery is low or internal malfunctions are detected). These autonomous features minimize the cognitive load on the person and reduce the risk of losing or damaging the UAV.

While partially autonomous UAVs are suitable in controlled settings, such as filming sports activities, fully autonomous behavior is suitable in situations where external controls cannot be easily communicated. For instance, autonomous mission planning is essential for such applications as remote surveillance and rescue operations, aiding police in locating and following a fleeing suspect, etc.

3. State-of-the-art approaches

Perception, planning, control, and interaction are the major computational components of an autonomous person-following robotic system. This section discusses these components of the state-of-the-art methodologies and their underlying algorithms.

3.1. Perception

An essential task of a person-following robot is to perceive the relative position of the person in its operating environment. The state-of-the-art perception techniques for object-following or object-tracking can, in general, be categorized based on two perspectives: feature perspective and model perspective (see Figure 5). Based on whether or not any prior knowledge about the appearance or motion of the target is used, the techniques can be categorized as model-based or model-free. Conversely, based on the algorithmic usage of the input features, perception techniques can be categorized as feature-based tracking, feature-based learning, and feature or representation learning.

Our discussion is schematized based on the feature perspective, since this is more relevant to the person-following algorithms. Additionally, various aspects of using human appearance and motion models are included in our discussion. These aspects, including other operational details of the state-of-the-art perception techniques for ground, underwater, and aerial scenarios, are presented in the following sections.

3.1.1. Ground scenario. The UGVs navigate in a two-dimensional (2D) space while following a person (Figure 6). Most UGVs adopt a unicycle model (Pucci et al., 2013) with linear motion along the ground (XY) plane and angular motion about the vertical (Z) axis. One implicit assumption is that the cameras are static and rigidly attached to the robots, as omnidirectional cameras (Kobilarov et al., 2006) are rarely used for person-following applications. The camera feeds and other sensory inputs are fused and sent to the perception module in order to localize the person with respect to the robot. Although the underlying algorithms vary depending on the choice of sensors, they can be generalized into the following categories.

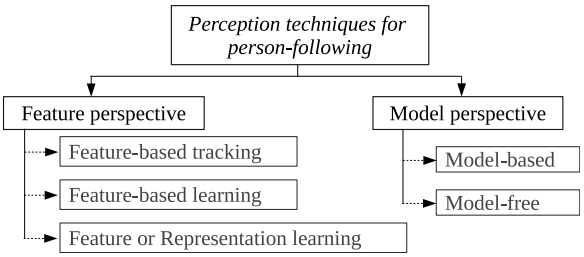
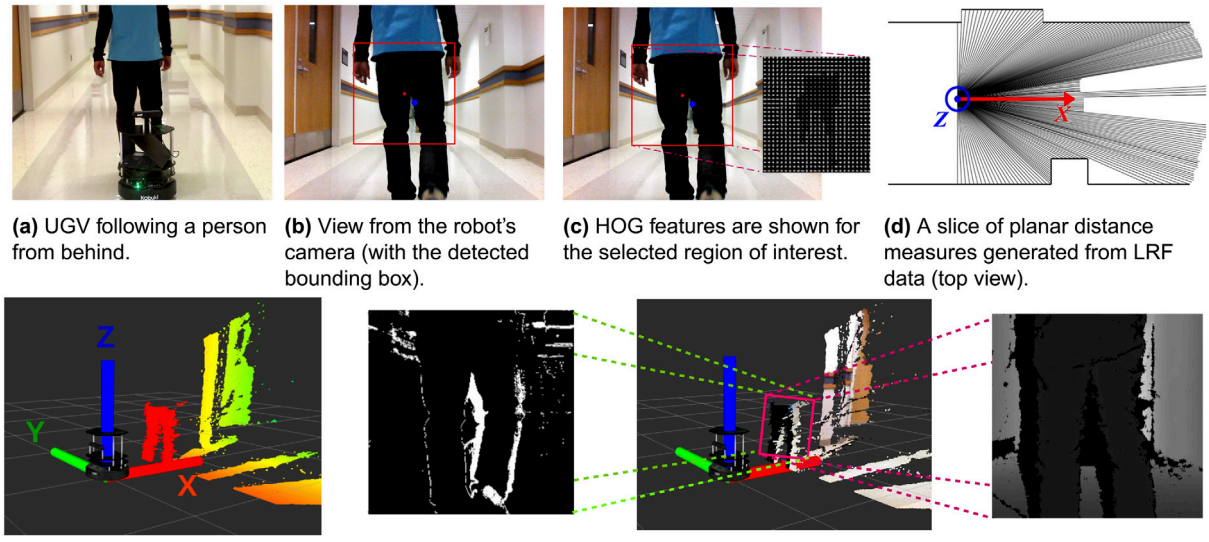


Fig. 5. Categorization of various perception techniques for person-following based on *feature* and *model perspectives*.

(i) Feature-based tracking. The simplest class of person-following algorithms detect person-specific features in the input feature space. For example, blob detection algorithms use color-based segmentation to track a person in the RGB image space (Hu et al., 2007, 2009; Schlegel et al., 1998). The obvious dependency on specific colors (e.g.,the person's clothing), make these algorithms impractical for general applications. More robust and portable algorithms can be designed using the depth data generated from an RGBD camera or a stereo camera. As illustrated in Figure 6(e), the presence of a person corresponds to a specific pattern in terms of shape, average distance, and the number of points in the 3D point-cloud. Usually, a template is designed based on the expected values of these attributes, and is then used for detection (Isobe et al., 2014; Satake et al., 2013). A family of person-following algorithms applies similar methodologies to LRF and sonar data. As seen in Figure 6(d), slices of planar distance measures from an LRF or directional distance measures from a sonar can be used to detect specific feature patterns or templates pertaining to a person in an upright posture.

More advanced algorithms iteratively refine the initial detection of person-specific features. Mean-shift and particle filter-based algorithms (Germa et al., 2009; Kwolek, 2004) are the most popular ones used for person-following. A mean-shift algorithm performs back-projection to find the probabilities of the target feature map in each point in the feature space. Then, by iteratively following the center of mass of the probability distribution (termed the mean-shift vector), the algorithm finds the mode of the distribution that corresponds to the best match for the target feature map. These approaches work very well for unimodal cases and are therefore not very effective in tracking several targets at once. Particle filters, however, adopt an iterative prediction-update process to derive a set of particles (i.e., candidate solutions). The particles are initialized randomly over the feature space, then iteratively updated based on their similarities with the



(a) UGV following a person from behind.

(b) View from the robot's camera (with the detected bounding box).

(c) HOG features are shown for the selected region of interest.

(d) A slice of planar distance measures generated from LRF data (top view).

(e) 3D point clouds for the global scenario are shown in the leftmost image; for a selected region within the robot's field of view, the background-subtracted binary image and the corresponding depth map are shown on the right.

Fig. 6. Snapshot of a person-following scenario by a UGV, the sensory data for different choices of sensors, and visualizations of the processed data used by various algorithms.
HOG: histogram of oriented gradients; LRF: laser range finder; UGV: unmanned ground vehicle.

target feature map. The update rules and similarity functions are designed in such a way that the particles move toward more prospective regions in the feature space and eventually converge to the target region.

Since searching over the entire feature space can be computationally expensive, it is very helpful to use prior knowledge or to make educated assumptions in order to reduce the search space. For example, Gaussian mixture model-based background subtraction (Stauffer and Grimson, 1999) can be used to avoid searching for person-specific features over background regions (Figure 6(e)). Additionally, exhaustive searching in every frame can be avoided by keeping track of the detected features over sequences of frames. Optical flow-based methods (Handa et al., 2008) and other feature-based trackers (Satake et al., 2012, 2013) take advantage of this for efficient tracking. Furthermore, educated assumptions of the walking model of a person can also facilitate the removal of unpromising regions from the feature space (Guevara et al., 2016).

(ii) Feature-based learning. Another class of approaches makes statistical assumptions about the true underlying function that relates the input feature space to the exact location of the person and then uses machine learning techniques to approximate that function. For example, histogram of oriented gradients (HOG) features are used to train support vector machines (SVMs) for robust person detection (Satake and Miura, 2009). HOG features are histograms of local gradients over uniformly spaced rectangular blocks in the image space. The localized gradient orientations are binned to generate dense feature descriptors. These descriptors, along with other sensory inputs (e.g., depth information) are used to formulate the feature space, which is then used for offline training of detectors such as SVMs. These detectors are known to be robust and their inference is fast enough for real-time applications. Other supervised models, such as decision tree and logistic regression, can also be applied by following a similar methodology (Germa et al., 2009). Figure 6(c) shows HOG features for a particular image patch; as seen, the presence of a person results in a particular spatial pattern in the HOG feature space.

Conversely, learning models based on adaptive boosting (AdaBoost) (Chen et al., 2017b) are different in that, instead of learning a single hypothesis, they iteratively refine a set of *weak* hypotheses to approximate the *strong (optimal)* hypothesis. The use of a number of learners almost always provides better performance than a single model in practice, particularly when the input features are generated using heterogeneous transformations (e.g., linear and non-linear) of a single set of inputs or simply contain data from different sources (i.e., sensors). Dollár et al. (2009) exploited this idea to extract *integral channel features* using various transformations of the input image. Such features as local sums, histograms, Haar features (Papageorgiou et al., 1998) and their various generalizations are efficiently computed using integral images and

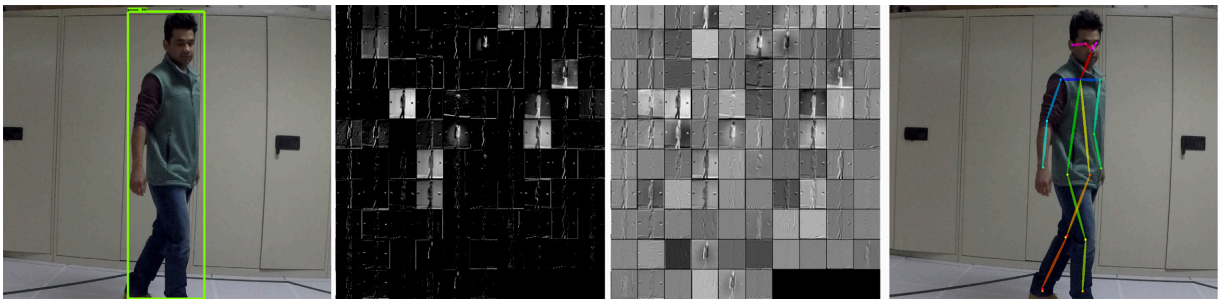


Fig. 7. The leftmost image shows a person detected by a deep object detection model, named Single-Shot Multibox Detector (Liu et al., 2016); visualizations for different feature representations that are extracted at the first two layers of the model are shown in the next two images; the rightmost image shows the human body pose being detected using Open-Pose (Cao et al., 2017).

then used as inputs to decision trees that are then trained via AdaBoost. A family of these models (Dollár et al., 2010; Zhu et al., 2006) is known to work particularly well as pedestrian detectors for near real-time applications, such as person-following.

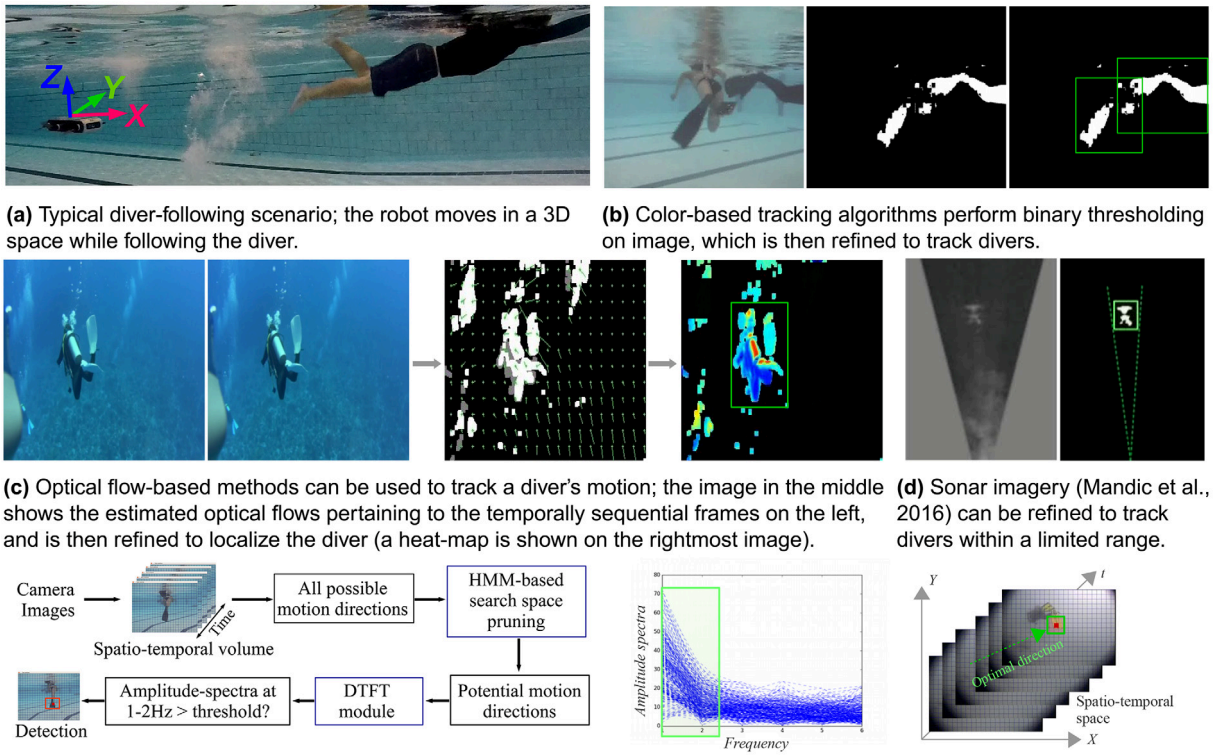
Furthermore, Bayesian estimation and other probabilistic models (Alvarez-Santos et al., 2012; Guevara et al., 2016) are widely used to design efficient person detectors. These models make statistical assumptions about the underlying probability distributions of the feature space and use optimization techniques to find the optimal hypothesis that maximizes the likelihood or the posterior probability. A major advantage of these models is that they are computationally fast and hence suitable for on-board implementations.

(iii) Feature or representation learning. Feature-based learning methods learn an optimal hypothesis on a feature space that is designed beforehand from the input sensory data. Consequently, the performance of the underlying algorithm largely depends on how discriminative and informative the feature space is. Deep learning-based approaches try to learn an optimal feature space and the optimal hypothesis simultaneously, providing a significant boost in performance. Recent developments in convolutional neural networks (CNNs) have made it possible to use these models in real-time applications such as person-following.

Supervised deep models typically consist of a convolutional network and an additional classification or regression network. The former consists of several convolutional layers that extract the informative features from the input data to generate different feature representations. These feature representations are then fed to a classification or regression network (a set of fully connected layers) for detection. Often, a separate region proposal network is used to allow efficient detection of several objects in the scene. Back-propagation and gradient-based optimization techniques are used to find the optimal feature space and optimal hypothesis simultaneously. The sample DeConvnet visualization (Zeiler and Fergus, 2014) shown in Figure 7 shows feature representations extracted by a CNN at different layers. Each sub-plot represents the feature maps that excite a particular neuron in the given layer. Feature maps for the first and second layers (of a CNN) are shown, since they are easier to inspect. These feature maps are used by the classifiers and regressors to detect a person and other objects in the scene.

The CNN-based deep models define the current state of the art for object detection, classification, and visual perception in general (Tensorflow, 2017). However, they require a set of comprehensive training samples in order to achieve good generalization performances by avoiding over-fitting. Nevertheless, they often perform poorly in such situations such as occlusions, appearance changes of the target (person), or random changes in the environment. Online learning schemes (Chen et al., 2017a) can cope with these issues by adjusting their model weights based on new observations on the fly. Conversely, in deep reinforcement learning (Chen et al., 2017e) and agent-based models (Gascueña and Fernández-Caballero, 2011), a reward function is defined based on the robot's perceived *state* and performed *actions*. The robot learns sequential decision making to accumulate more rewards while in operation. The overall problem is typically formulated as a Markov decision process and the optimal action-state rules are learned using dynamic programming techniques. These methods are attractive because they do not require supervision and they imitate the natural human learning experience. However, they require complex and lengthy learning processes.

Unsurprisingly, modern person-following robots use deep learning-based person detectors (Chen et al., 2017a; Jiang et al., 2018; Wang et al., 2018b) since they are highly accurate in practice and robust to noise, illumination changes, and other visual distortions. More advanced robots go beyond person detection and adopt robust models for human pose estimation and action recognition. These are potentially useful for enabling many additional capabilities, such as learning long-term human behavior, understanding sentiment, or engaging in natural conversation;



(a) Typical diver-following scenario; the robot moves in a 3D space while following the diver.

(b) Color-based tracking algorithms perform binary thresholding on image, which is then refined to track divers.

(c) Optical flow-based methods can be used to track a diver's motion; the image in the middle shows the estimated optical flows pertaining to the temporally sequential frames on the left, and is then refined to localize the diver (a heat-map is shown on the rightmost image).

(d) Sonar imagery (Mandic et al., 2016) can be refined to track divers within a limited range.

(e) Mixed-domain periodic motion tracker (Islam and Sattar, 2017) uses a HMM to track the potential motion directions of a diver, which are then validated using frequency responses of the corresponding motion; as the right plot shows, intensity variations in the spatio-temporal domain along diver's swimming directions correspond to high-energy responses of 1–2 Hz in the frequency domain.

(f) A convolutional neural network based deep model is used to detect divers and other objects (e.g., other robots) by an autonomous underwater vehicle (Islam et al., 2018a).

Fig. 8. Snapshots of autonomous diver-following scenarios and visualization of the processed data used by various algorithms. DTFT: discrete-time Fourier transform; HMM: hidden Markov model.

these are attractive for interactive person-following applications in social settings.

3.1.2. Underwater scenario. As discussed in Section 2.2, perception is more challenging for underwater diver-following robots. Challenging operating conditions call for two major characteristics of a perception algorithm: robustness to noisy sensory data and fast running time with limited on-board resources. Consequently, state-of-the-art approaches focus more on robustness and fast running time than on accuracy of perception.

To this end, simple feature-based trackers are often practical choices (Sattar and Dudek, 2006). As illustrated in Figure 8(b), color-based tracking algorithms can be utilized to localize a diver in the image space. These algorithms perform binary image thresholding based on the color of the diver's flippers or suit. The binary image is then refined to track the centroid of the target (diver) using algorithms, such as mean-shift, particle filters, etc. AdaBoost is another popular method for diver tracking (Sattar and Dudek, 2009a); as discussed in Section 3.1.1, AdaBoost learns a strong tracker from a large number of simple feature-based trackers. Such ensemble methods are proven to be computationally inexpensive yet highly accurate in practice. Optical flow-based methods can also be utilized to track a diver's motion from one image frame to another, as illustrated in Figure 8(c). Optical flow is typically measured between two temporally ordered frames using the Horn and Schunk formulation (Inoue et al., 1992), which is driven by brightness and smoothness assumptions of the image derivatives. Therefore, as long as the target motion is spatially and temporally smooth, optical flow vectors can be reliably

used for detection. Several other feature-based tracking algorithms and machine learning techniques have been investigated for diver tracking, and underwater object tracking in general. However, these methods are mostly applicable in favorable visual conditions with clear visibility and steady lighting.

Color distortions and low visibility issues are common in deep-water scenarios. Sattar and Dudek (2009b) showed that human swimming cues in the frequency domain are more stable and regular in noisy conditions than traditionally used spatial features like shape and color. Specifically, intensity variations in the spatio-temporal domain along the diver's swimming direction have identifiable signatures in the frequency domain. These intensity variations caused by the diver's swimming gait tend to generate high-energy responses in the 1–2 Hz frequency range. This inherent periodicity can be used for robust detection of divers in noisy conditions. A mixed-domain periodic motion (MDPM) tracker generalizes this idea in order to track arbitrary motions (Islam and Sattar, 2017). In the MDPM tracker, spatial features are used to keep track of the potential motion directions of the diver using a hidden Markov model (HMM). Frequency responses along those directions are then inspected to find the most probable one; the overall process is outlined in Figure 8(e). These methods are fast and known to be more robust than simple feature-based trackers.

The use of sonars is effective in unfavorable visual conditions. Sonars use acoustic chirps (low-frequency sound waves) along a range of bearings; directional distance measures are then calculated from the reflected sound waves. AUVs and autonomous surface vehicles (ASVs) most commonly use *active* sonars to track divers (Mandic et al., 2016; Miskovic et al., 2015) in diver-following applications. Additionally, the processed sonar image measurements (Figure 8(d)) can be fused with USBL measurements to obtain reliable tracking estimates at a steady rate. Such sensor fusion increases robustness and works even in cases when either the sonar or the USBL measurements are noisy or unavailable. However, active sonars face challenges in coastal waters, owing to scattering and reverberation. Additionally, their usage cause disturbances to the marine ecosystem and may also be limited by government regulations on sound levels. Thus, the use of passive sonars such as hydrophones is a practical alternative (Gemba et al., 2014; Hari et al., 2015). Passive sonars capture sound waves of diver's breaths and movements, which have inherent periodicity. These waves are then analyzed in the frequency domain to detect periodic bursts of low-frequency sound waves pertaining to the diver's breathing or movements. A similar methodology is used by underwater ROVs that use electric-field sensors to detect the presence of divers within a short range (Lennartsson et al., 2009).

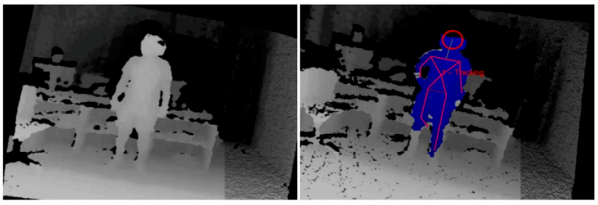
Deep learning-based object detection models have recently been investigated for underwater applications as well (Islam et al., 2018a; Shkurti et al., 2012). The state-of-the-art pre-trained models are typically trained (offline) on large underwater datasets and sometimes quantized or pruned in order to get faster inference by balancing robustness and efficiency (Islam et al., 2018a,c). As illustrated in Figure 8(f), once trained with sufficient data, these models are robust to noise and color distortions; additionally, a single model can be used to detect (and track) several objects at once. Despite the robust performance, these models are not as widely used in practice as in terrestrial scenarios, owing to their slow on-board running times. However, with the advent of mobile supercomputers and embedded parallel computing solutions (Google, 2018; NVIDIA, 2014), efficient on-board implementations of these models are becoming possible. Nevertheless, although the online learning and reinforcement learning-based approaches are effective for person tracking when the appearance and scene changes (Chen et al., 2017a,e), they are yet to be successfully used in practice for diver-following applications.

3.1.3. Aerial scenario. The underlying mechanism of a perception algorithm for person-following UAVs is mostly defined by two aspects: the expected flying trajectory of the UAV and the available sensory data. For instance, in some personal applications, the UAV flies close to the person at a low altitude (e.g., 4–6 m from the ground). The perception algorithms in such a set-up are similar to those in ground scenarios, as illustrated in Figure 9(a). Conversely, applications such as filming sports activities demand rather complex trajectories of the UAV while following an athlete (Figure 9(d)). Robust detection and control mechanisms are required in these set-ups, including reliable sensing. Lastly, autonomous surveillance and rescue operations involve sensing from long distances, often in unfavorable sensing conditions; hence, perception techniques differ among these scenarios.

Standard feature-based tracking methods are suitable if the UAV is expected to fly close to the person while maintaining a smooth horizontal trajectory. As seen in Figure 9(a) and (b), the camera image captures most of the person's body with no significant perspective distortions; hence, computer vision-based object detectors and pedestrian detectors perform well in such scenarios. To this end, color-based segmentation, mean-shift, particle tracker, and HOG-based detectors are widely used (Higuchi et al., 2011; Kumar et al., 2011; Lugo and Zell, 2014; Teuliere et al., 2011). The operations of these algorithms are discussed in Section 3.1.1. In a seminal work, Pestana et al. (2014) showed that OpenTLD trackers can achieve robust performance for outdoor suburb environments. These trackers decompose a target-tracking task into tracking, learning, and detection (TLD); they are known to perform well for long-term object tracking in general. One limitation of TLD trackers is that they often incorporate the background into the learning over time, leading to quick target drift.

More reliable sensing and robust tracking performance can be achieved by using an additional depth sensor (e.g., a

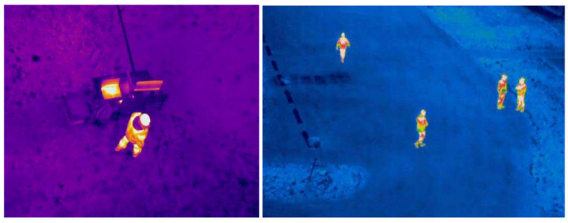
(a) Typical aerial person-following scenario; the UAV moves in 3D space while following the person. The rightmost figure shows the target bounding boxes found by the mean-shift algorithm on the color thresholded binary image; the final target bounding box (marked in pink) is projected on the original image and refined for tracking.



(b) A person being tracked from stabilized depth-map using OpenNI skeleton tracker (Naseer et al., 2013).

(c) Illustration of different appearances of humans based on the altitude of UAVs; a low-altitude case is shown on the left and a higher-altitude case on the right (Mueller et al., 2016).

(d) Illustrations of commercial UAVs following an athlete: (left) by creating a 3D map of the surrounding using visual SLAM & tracking him within the map (Skydio, 2018); (right) by tracking a paired wrist-mounted device containing a GPS receiver (Staaker, 2016).



(e) Snapshots taken by a military-grade thermal infrared camera (SPi, 2015); these images capture the infrared energy reflected by objects (such as humans), which can be used to isolate them from the background.

Fig. 9. Snapshots of aerial person-following scenarios by UAVs and visualization of the processed sensory data used by various algorithms.
GPS: Global Positioning System; SLAM: simultaneous localization and mapping; UAV: unmanned aerial vehicle.

RGBD camera), particularly for indoor applications. Naseer et al. (2013) presented an indoor person-following system using two cameras; a regular camera for determining the 3D position of the UAV based on markers on the ceiling and a depth camera to detect a person in 3D. The images from the depth camera are warped based on the calculated 3D position. The stabilized depth images are then used for robust perception using the OpenNI platform (Figure 9(b)). Gioioso et al. (2014) also used an RGBD camera to detect hand-gesture-based teleoperation commands for UAVs. Such systems, however, are limited to indoor environments and small motions. Additionally, they often require a remote computer for intensive computations.

For challenging outdoor applications, where the UAV trajectory changes rapidly because of dynamic obstacles or fast movements, the person may appear significantly different from different viewpoints. Hence, perspective distortions need to be taken into account. De Smedt et al. (2015) used *ground plane estimation* techniques to approximate the orientation of the ground plane in 3D relative to the position of the camera; object heights in the image were then predicted based on the homography of the ground plane and the real-world sizes of the objects. De Smedt et al. (2015) exploited this idea to localize prospective rectangular regions in the image space for detecting pedestrians of expected heights between 160 cm and 185 cm. This allows approximation of the height of the person in different perspectives and thus reduces the search space, leading to efficient tracking performances. De Smedt et al. (2015) used standard pedestrian trackers based on aggregate channel features and achieved good on-board performances. A number of other online tracking algorithms have been investigated by Mueller et al. (2016) for person-following and general object tracking by UAVs. Mueller et al. (2016) also presented a *camera handover* technique, where one UAV can pass the tracking task over to another UAV without interruption; this can be useful in long-term tracking and filming applications. Moreover, some commercial UAVs build a 3D map of the surroundings using techniques such as visual simultaneous localization and mapping (SLAM) and follow their target (person) within the map (Skydio, 2018). These UAVs are usually equipped with advanced features, such as obstacle avoidance or target motion prediction. Furthermore, UAVs that capture sports

activities often track the location information provided by a paired controller device carried by or mounted on an athlete, as illustrated in Figure 9(d). The paired device is equipped with a GPS receiver and communicates information related to motion and orientation of the athlete; this additional information helps the UAV plan its optimal trajectory for filming (Vasconcelos and Vasconcelos, 2016).

As mentioned, thermal and infrared cameras are particularly useful in autonomous surveillance, search and rescue, and other military applications. Thermal images detect the emissions of heat from various objects in the image space, which can be easily identifiable from the surroundings. This feature is crucial while sensing from long distances and in unfavorable sensing conditions. Figure 9(e) shows snapshots of a military-grade thermal infrared camera (SPi, 2015); as seen, the warm objects can be easily located in the image. In fact, one major advantage of using thermal imagery is that simple computer vision techniques can be used for robust detection. For instance, Portmann et al. (2014) showed that the standard background subtraction techniques can be used to segment regions that are both hotter and colder than the environment. Then, HOG-based detectors or particle filters can be used to track humans in the segmented regions. Gaszczak et al. (2011) used the mean-shift algorithm on the background-subtracted thermal image and achieved good results. Additionally, they showed that the Haar classifiers can be used to detect human body signatures, as well as other objects, accurately. These methods are computationally inexpensive and suitable for on-board implementations.

Lastly, deep learning-based person detectors are yet to be explored in depth for the aerial applications, largely owing to the limited on-board computational resources available, particularly in consumer-grade UAVs. Nevertheless, some recent commercial UAVs, such as Skydio R1 (Skydio, 2018), use mobile embedded supercomputers in order to use deep visual models in real time. It is safe to say that with faster mobile supercomputers and better low-power computing solutions, efficient on-board implementations of various deep learning-based perception modules will soon become possible and will be more commonly used by person-following UAVs in the near future.

3.2. Planning and control

Once the perception module obtains an estimate of the target (person) pose by processing the sensory inputs, control commands need to be generated in order to achieve the desired motion. Ground robots navigate in a 2D plane, whereas underwater and aerial robots navigate in 3D spaces; hence, the corresponding control signals and their operational constraints vary. However, the overall operation is mostly similar; Figure 10 illustrates an outline of the operational flow for an autonomous person-following system. The following discussions provide an overview of the planning and control modules that are standard for general object-following, and focus on the aspects that are particularly important for person-following applications.

First, the target person's position and heading information are estimated with respect to the robot's known frame of reference. Additionally, the sensory data are processed and sent to the state estimation filters. These *observed measurements* are used by the filters to refine the state estimation through iterative *prediction-update* rules. Linear quadratic estimators, such as the Kalman filter (Kalman, 1960), and non-linear estimators, such as the extended Kalman filter (EKF) (Julier and Uhlmann, 1997) are most widely used for this purpose. The unscented Kalman filter (Wan and Van Der Merwe, 2000) addresses the approximation issues of the EKF and is often used in practice for state estimation from noisy sensory data. Methodological details of these algorithms are beyond the scope of this paper; interested readers are referred to Jung and Yi (2012); Lugo and Zell (2014); Morioka et al. (2012); Satake and Miura (2009); Sattar et al. (2008); Teuliere et al. (2011) and Yoon et al. (2014).

The refined measurements are then processed to generate a set of *way points*, i.e., a representation of potential trajectories for the robot in order to follow the target person. The path planner uses this information and finds the optimal trajectory by taking into account the estimated relative positions of the static obstacles, other humans, and dynamic objects in the environment. The constraint here is to optimize some aspect of the anticipated motion of the robot, such as travel time, safety, or smoothness of motion. A sequence of points is then generated to discretize the anticipated motion; the points pertaining to the optimal trajectory are generally termed the *set-points* (Doisy et al., 2012). Finally, the control modules analyze the set-points and generate navigation commands to drive the robot. The generated navigation commands are usually fed to a set of feedback (e.g., proportional–integral–derivative (PID)) controllers. This process of robot control is also generic for most applications; interested readers are referred to De Wit et al. (2012), Mezouar and Chaumette (2002), and Pounds et al. (2010) for further details.

Figure 11 illustrates a categorical overview of various types of path planning algorithm in the literature. Based on the locality of sensing, planning can be either *global* (for fully observable environments) or *local* (for partially observable environments). Additionally, if the optimal path (to the target) is computed first and executed sequentially, it is termed as *offline* planning; conversely, the planned path is refined dynamically in *online* planning paradigms. Since person-following robots are deployed in partially observable and dynamic environments, they require local and online path planning in order to adapt to irregular and unpredictable changes in their surroundings. Path planning algorithms for person-following robots can be further categorized based on mapping information and on their algorithmic structure. Although these algorithms are fairly standard for dynamic target-following, a brief discussion of

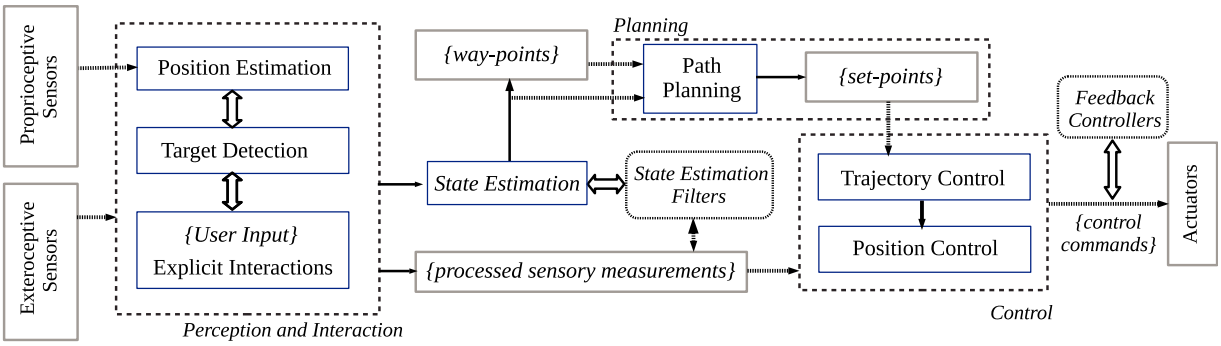


Fig. 10. Data and control flow among major computational components of a person-following system; this flow of operation is generic to most autonomous object-following paradigms.

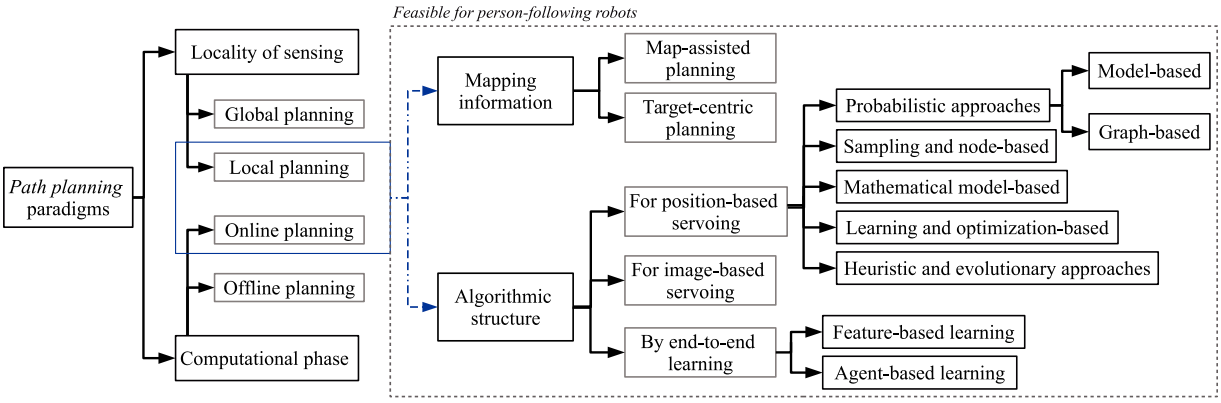


Fig. 11. Categorization of path planning algorithms from the perspective of sensing, methodology, and computation.

their operational considerations for person-following is presented in the following sections.

3.2.1 Map-assisted versus target-centric planning. Map-assisted planning is feasible for structured environments with a known map, particularly for person-following ground robots in indoor settings (Doisy et al., 2012; Nikdel et al., 2018). The global map of the environment (including static objects) is used as prior knowledge. Typically, a *static planner* keeps track of the robot's location within the map and its admissible waypoints by taking into account static obstacles in the environment (Ahn et al., 2018). The *dynamic planner* then refines these waypoints by considering the motions of the dynamic objects in the environment (Figure 12). Other constraints, such as social awareness, implicit interactions, etc., can also be considered in the refinement process that eventually generates the optimal path (Cosgun et al., 2016). Standard map-based navigation techniques are then used to invoke the person-following motion.

Although a global map can significantly simplify the planning and navigation processes, it is usually not available in outdoor applications. In such cases, a target-centric approach is adopted. First, the locally sensed information is used to create a partial map of the environment; traditional SLAM-based techniques are most commonly used for this purpose (Ahn et al., 2018; Cosgun et al., 2016). As illustrated in Figure 13, the UAV creates a 3D (depth) map of the partially observed environment in order to find the optimal path for person-following (Skydio, 2018). Such *reactive* path planning sometimes leads to non-smooth trajectories, particularly if the person moves quickly in a zigzag trajectory (Tarokh and Merloti, 2010). Anticipatory planning, i.e., predicting where the person is going to be next and planning accordingly, can significantly alleviate this problem and is thus widely used in practical applications (Hoeller et al., 2007; Nikdel et al., 2018; Tarokh and Shenoy, 2014).

3.2.2 Planning for position-based servoing. In position-based servoing, the path planner finds the optimal path to follow a target using its estimated position with respect to the robot. For instance, a person-following UAV uses its current 3D position as the *source* and the estimated 3D location of the person as the *destination*, and then uses source-to-destination path planning algorithms to find the optimal path that meets all the operational constraints. It is to be noted that this planning can be either map-assisted or target-centric, depending on whether or not global mapping information is available.

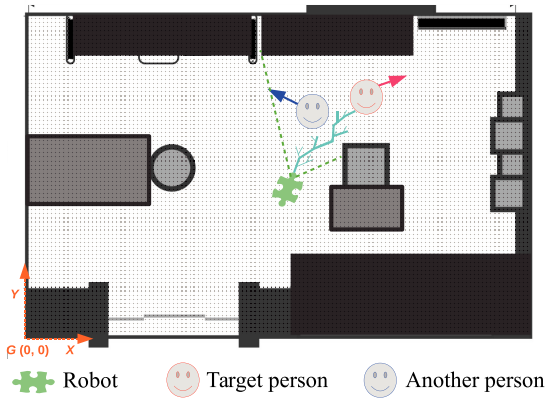


Fig. 12. Map-assisted 2D path planning of ground robot, avoiding static and dynamic obstacles within the map.

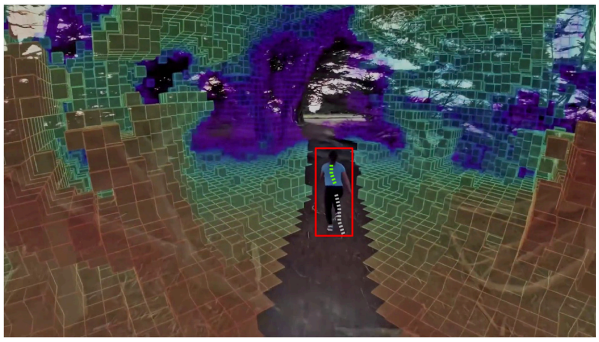


Fig. 13. An unmanned aerial vehicle using a multi-camera depth map of the partially observed environment for target- centric planning in order to follow a person (Skydio, 2018).

Standard path planners typically represent the state space using cells, grids, or potential fields and then apply various search methods to find the optimal source-to-destination path. For instance, the navigation space (and the locations of relevant objects) is often interpreted using an *occupancy grid*, and graph search-based algorithms, such as A*, D*, or IDA* (Iterative Deepening A*), are used to find the optimal path (Ahn et al., 2018; Huskić et al., 2017; Müller et al., 2008). Another approach is to randomly sample the state space and attempt to establish source-to-destination connectivity using such techniques as rapidly exploring random trees (RRTs) (Triebel et al., 2016), RRT*, or probabilistic road maps (Hoeller et al., 2007). These methods are good at finding near-optimal solutions at a fast rate in large search spaces where ensuring completeness is computationally expensive; hence, they are widely used in real-time path planning for person-following robots.

It is also common to represent the planning hypothesis given the constraints as a probabilistic inference model. Then, the problem reduces to finding a minimum cost or maximum utility path from the search space of all admissible paths. Machine learning models and heuristic and evolutionary approaches are also used to approximate the optimal solution (i.e., to find a near-optimal path), particularly if the search space is too large (Gong et al., 2011; Triebel et al., 2016). Moreover, the problem can be modeled as a partially observable Markov decision process (POMDP) in order to perform online planning in a continuous state and action space (Goldhoorn et al., 2014; Triebel et al., 2016). POMDPs are good at dealing with dynamic environments and complex agent behaviors. However, they can be computationally intractable and generate sub-optimal solutions. Therefore, approximate solutions are typically formulated with an assumption of a discrete state or action space.

Table 5 summarizes the different classes of path planning algorithms for position-based servoing and highlights their operational considerations in person-following applications. These algorithms are fairly standard; interested readers are referred to González et al. (2015) and Yang et al. (2016) for further methodological details.

3.2.3. Planning for image-based servoing. Autonomous navigation of a robot using visual feedback is known as image-based (visual) servoing, where the path planner uses image-based features in order to find the optimal path to follow the target (Gupta et al., 2017). Image-based servoing is particularly useful when it is difficult to accurately localize the target, particularly underwater and in GPS-denied environments (Pestana et al., 2014). For instance, AUVs often use *bounding-box reactive* path planners for diver-following (Islam et al., 2018a). Here, the planning objective is to keep the target diver at the center of the robot's view. That is, motion commands are generated in order to bring the observed bounding box of the diver to the center of the camera image (Figure 14). The relative distance of the diver is approximated by the size of the bounding box and forward velocity rates are generated accordingly. In addition, the yaw and pitch commands are normalized based on the horizontal and vertical displacements of the observed bounding box center from the image center; these navigation commands are then regulated by the controller to drive the robot.

Furthermore, it is common to simplify the planning component for image-based servoing in order to reduce computational burdens on the robot. For instance, diver-following robots sometimes plan a straight-line trajectory to remain immediately behind the diver (Islam and Sattar, 2017). A similar strategy is adopted by ground robots as well (Brookshire, 2010; Doisy et al., 2012; Wang et al., 2018a), with an additional planning component for obstacle avoidance. As illustrated in Figure 15, person-following UGVs can use tools from prospective geometry to get the relative homography of the orthogonal planes and then find the optimal path along the ground plane by keeping safe distances from the person and obstacles. This simplifies the operational complexities and is often sufficient for non-critical applications.

Table 5. Various classes of path planning algorithms used by person-following robots for position-based servoing (based on the categorization shown in Figure 11), and their operational consideration.

Categories	Operation	Constraints	Advantages	Selected references
Probabilistic approaches	The planning hypothesis given the constraints is represented as a generative or inference model	Probabilistic assumptions on the system model might not always hold	Computationally fast and good for planning with limited sensing	Park and Kuipers (2013); Sung and Chung (2016)
Sampling and node-based	The workspace is sampled into nodes, cells, grids, or potential fields; then different search methods are used to find the optimal path	Prior knowledge about the environment is needed	Good at finding sub-optimal and often optimal solutions	Hoeller et al. (2007); Huskić et al. (2017); Triebel et al. (2016)
Mathematical model-based	The environment is modeled as a time-variant kino-dynamic system; then a minimum cost path or maximum utility path is computed	Can be computationally expensive; analytic solution of the system may not exist	Reliable and optimal	Cosgun et al. (2013); Doisy et al. (2012); Tisdale et al. (2009)
Learning and optimization-based	Parameters of the optimal planning hypothesis are approximated using machine learning models	Requires accurate feature representation and rigorous training process	Can easily deal with complex and dynamic environments	Morales et al. (2004); Triebel et al. (2016)
Heuristic and evolutionary approaches	Planning hypothesis is evaluated using a heuristic or bio-inspired objective function; then an optimal solution is iteratively sought	Search space can be very large; often produce locally optimal solution	Good at dealing with complex and dynamic environments	Gong et al. (2011); Sedighi et al. (2004)

Fig. 14. Illustration of a bounding-box reactive planner; the horizontal and vertical displacements of the center of the detected bounding box is used for image-based servoing.

3.2.4. Planning by end-to-end learning. End-to-end techniques try to *learn* problem-specific robot navigation rules directly from input sensory data. This way of coupling the perception, planning, and control modules together is inspired by the *self-driving-car* concept and is very popular these days. Several deep-learning-based models for supervised learning and agent-based reinforcement learning have recently been introduced for person-following as well (Dewantara and Miura, 2016; Goldhoorn et al., 2014). Typically, these models are first trained in simulations with an existing motion planner and then transferred to real-world environments for further tuning. Researchers have reported exciting results, demonstrating their effectiveness for UGVs in autonomous 2D navigation, avoiding obstacles, following people in near-optimal paths, multi-agent cooperation, etc. However, these techniques are mostly applied for person-following UGVs in indoor settings, and sometimes only in simulation environments (Dewantara and Miura, 2016; Pierre, 2018). Therefore, more research attention is needed in order to improve and generalize these techniques for a wide range of other person-following applications.

3.2.5. Other considerations for planning and control. In addition to the operating constraints for person-following mentioned already, there are other practical, and often application-specific, considerations for effective planning and control. Several such aspects are discussed in this section.

Planning ahead to avoid occlusion. The most essential feature of a person-following robot's planning module is to ensure that the target person is in the field of view during the robot's motion. The trajectory needs to be planned in such a way that, in addition to meeting the standard criteria of an optimal path, e.g., based on distances from obstacles, expected travel time, smoothness of anticipated motion, etc., the person remains reasonably close to the center of the robot's field of view and not occluded by obstacles.

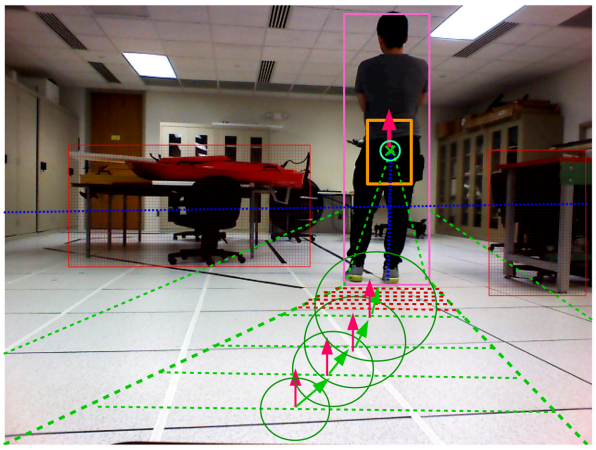


Fig. 15. Simple planning strategy for an unmanned ground vehicle; it finds a straight-line trajectory in order to remain immediately behind the person while avoiding obstacles.

This is challenging if the sensing range is limited, especially in the presence of dynamic obstacles. Typically, a probabilistic map (Hoeller et al., 2007) for motions of the moving objects in the scene is formulated and then path planning is performed on the dynamic occupancy field; a temporal window of *motion history* is maintained to facilitate such a formulation. Another approach is to predict the positions and velocities of the moving objects a few time-epochs into the future and plan the optimal trajectory ahead of time. Such anticipatory planning is particularly important for person-following UGVs that are meant to stay ahead of the person (Mi et al., 2016) and UAVs that film fast-moving athletes (Skydio, 2018).

Camera control. If the person-following robot is equipped with a pan-and-tilt camera, a controller module is required to ensure that the camera is always pointed at the person during navigation (Doisy et al., 2012; Tarokh and Merloti, 2010). In addition, it is common for person-following UAVs to have camera gimbals; if so, an additional module is required to control the gimbals' angles instantaneously (Skydio, 2018; Staaker, 2016) based on the person's relative motion.

Person re-identification and recovery. In addition to robust person detection and tracking, person-following robots need to be able to plan to re-identify when necessary (Koide and Miura, 2016). Moreover, these techniques are essential for accompanying a specific person (Eisenbach et al., 2015; Ilias et al., 2014). Predictive and probabilistic models, such as Kalman filters and particle filters, are typically used to estimate the person's future location, which can be used as prior knowledge in case of a *missing target* situation. That is, when the robot fails to detect the person (owing to occlusion or noisy sensing), the *recovery planner* can use that person's anticipated location as a prior and search probable locations for re-identification (Do Hoang et al., 2017; Gupta et al., 2017). Standard feature-based (Alvarez-Santos et al., 2012) and trajectory replication-based techniques (Chen et al., 2017a) are most commonly used in practice; appearance-based deep visual methods (Ahmed et al., 2015; Li et al., 2014) can also be used by recovery planners for person re-identification.

Additional planning and control procedures are required to incorporate desired autonomous behaviors in emergency situations, e.g., when the recovery planner fails to re-identify the missing person, or if there is a critically low battery or internal malfunctions are detected. For instance, a UAV should be capable of making an emergency landing and communicate its status to the person if possible. Moreover, UGVs and UAVs can use some sort of emergency beacons, e.g., flashing lights or beeping sounds, to attract immediate attention.

Social awareness in a crowded area. It is essential for person-following robots to maintain certain social rules while operating in a populated area (Honig et al., 2018). For instance, passing pedestrians on the correct side, maintaining average human walking speed, taking nearby persons' motions into account for planning, etc., are norms (Dewantara and Miura, 2016; Gockley et al., 2007) that a social robot should be aware of. Therefore, application-specific social norms need to be modeled and translated into path planning and control rules in order to enable the desired behaviors. Enabling socially aware behaviors in mobile robots is an active research topic and researchers have been exploring these aspects for person-following robots in various environments, such as airports (Triebel et al., 2016), hospitals (Ilias et al., 2014), and other crowded areas (Ferrer et al., 2013).

Maintaining the norms of interaction. Lastly, the planning and control modules for person-following robots need to accommodate the norms of explicit and implicit human–robot interactions. In particular, various aspects such as desired proximity behaviors, following angles (Shanee et al., 2016), turning and waiting behavior, etc., should be considered during trajectory planning. Additionally, application-specific choices, such as whether to stay behind or side-by-side while following, the desired speed, and relevant anticipative behaviors (Granata and Bidaud, 2012; Mi et al., 2016) are essential considerations. Another important feature is to maintain the expected behavior during explicit interactions (Hu et al., 2014; Islam et al., 2018c), e.g., being stationary when the human is communicating, and exhibiting correct acknowledgement responses. These aspects of human–robot interaction are elaborately discussed in the following section.

3.3. Interaction

Various forms of explicit and implicit interactions for person-following scenarios are discussed in Section 2. The following discussion provides a summary of how these

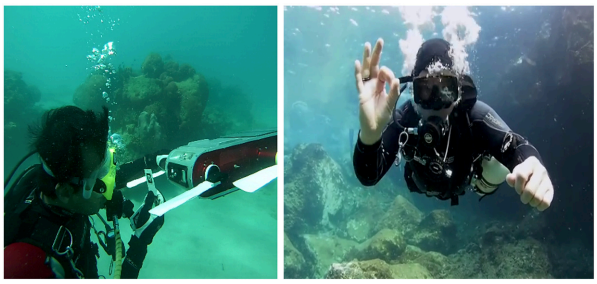
interactions take place, different methodologies used, and the related operational considerations.

3.3.1. Explicit interaction. Explicit interactions happen when there are direct communications between the human and the robot. Although most robots are equipped with peripheral devices and sometimes haptic sensors (Ghosh et al., 2014), those are typically used for offline tasks (powering the robot, transferring software or data, sending emergency signals, etc.). Conversely, communication paradigms based on speech, tags or markers, and hand gestures are used during operation for explicit human–robot interaction.

Verbal communication is convenient and commonly practiced in ground applications (Sakagami et al., 2002). Typically the person uses voice commands to convey simple instructions to the robot. The level of communication can vary from simple imperative commands (start or stop following, move left or right) to complex procedural instructions (e.g., a set of sequential tasks) depending on the application requirements. Systems for speech recognition and synthesis are very robust and commercially available these days. However, their usage is mostly limited to terrestrial applications.

Tags or fiducial markers (e.g., ARTag, ARToolkit) have been used for underwater human–robot communication. Visual languages, such as RoboChat (Dudek et al., 2007), assign different sequences of symbolic patterns of those markers to a set of grammar rules (Figure 16). This is generally a robust way of communication because the fiducial markers are easy to detect in noisy underwater conditions. However, it is not very intuitive, and carrying a large set of markers during an underwater mission is inconvenient. Consequently, hand-gesture-based communication paradigms (Chiarella et al., 2015; Islam et al., 2018b) are often preferred, where sequences of hand gestures are used as symboling patterns instead of the tags. Detecting hand gestures in real time is relatively more challenging than detecting markers; therefore, deep visual detectors are typically used to ensure the robustness and accuracy of the system (Islam et al., 2018c).

An additional challenge for hand gesture-based communication in aerial applications is the relatively long and varying human–robot distance (Figure 17). Unlike in an underwater scenario, the person cannot come close to the robot and perform hand gestures in front of its camera. Consequently, the UAV might end up being too far away to detect various kinds of hand gestures (Bruce et al., 2016; Monajjemi et al., 2016). In such cases, it is often useful to use a reliable gesture (a static palm gesture, waving hands, etc.) to instruct the UAV to first come closer and then perform other hand gestures for communication (Cauchard et al., 2015; Naseer et al., 2013). Conversely, hand gesture-based communication is relatively less challenging in ground applications (Alvarez-Santos et al., 2014; Marge et al., 2011) and sometimes used even if a voice-based communication system is available. Moreover, it is often more feasible than voice-based communication in crowded environments (Ferrer et al., 2013), and in a multi-robot setting.



(a) Using AR-tags. **(b)** Using hand gestures.

Fig. 16. Diver communicating instructions to autonomous underwater vehicle during a mission (Islam et al., 2018c).

Fig. 17. The visual challenges of detecting hand gestures from a distant unmanned aerial vehicle (Monajjemi et al., 2016): notice the minuscule appearance of people on the right compared with the left image where the UAV is much closer.

Smart devices and paired wearable devices are also commonly used to communicate human instructions, particularly by commercial UAVs (Skydio, 2018; Vasconcelos and Vasconcelos, 2016), and interactive UGVs (Burgard et al., 1998; Faria et al., 2015). Typically, the humans use a set of menu options to instruct the robot to perform specific tasks. For instance, instructions to start or stop recording videos, move in a particular direction, stop following, make an emergency landing, etc., are practically useful for interacting with person-following UAVs. Conversely, a conversational user interface is needed for UGVs that serve as museum tour-guide robots, or as personal assistants.

3.3.2. Implicit interaction. Research studies on implicit interactions in person-following scenarios mostly concentrate on two aspects: the human perspective and the robot perspective. As mentioned in Section 2, these aspects boil down to the following sets of mutual responsibilities in different stages of a person-following operation.

1. *Spatial conduct* consists of a set of desired proxemic behaviors (Fleishman et al., 2018) of a robot while following a person in a human–robot interaction setting.

This behavioral quantification is important to define safe person-following trajectories and to model proximity control parameters (Yamaoka et al., 2008) for following and waiting while engaging in, and during, an explicit interaction.

2. *Appearance and gaze conduct* consists of desired responsive behaviors during human–robot communication (Zender et al., 2007) and non-responsive gaze behaviors during following, approaching, and handing-over scenarios (Moon et al., 2014), etc. Moreover, for person-following robots that stay in front (Jung and Yi, 2012; Nikdel et al., 2018), companion humans' gaze behavior is an important feature to track in order to predict their motion. Additionally, it helps the robot to identify when the person is about to start an explicit interaction and to plan accordingly, e.g., slow down or stop, prepare to detect and interpret the communicated instructions, etc.
3. *Motion conduct* refers to a set of desired person-following motion trajectories of the robot in different situations. It includes motion models for following a person from different directions (from behind, side-by-side, at an angle, etc.), turning behaviors, and waiting behaviors. Additionally, the expected motion behavior of the robot when its human companion is interacting with other people or goes out of its sight are important design considerations (Gockley et al., 2007; Granata and Bidaud, 2012). Motion conduct rules are used by the planning component of the system in order to maintain the desired motion behaviors. Therefore, prior knowledge about human motion (walking, swimming, etc.) and the overall interaction model can facilitate the design of those anticipated motion behaviors (Hu et al., 2014) for person-following robots.

The modalities and characteristics of implicit interactions are difficult to quantify in terms of technicality. This calls for rigorous user studies and feasibility analysis to formulate the right research questions and their effective solutions.

4. Qualitative analysis: Feasibility, practicality, and design choices

An overwhelming amount of research work and industrial contributions have enriched the literature on autonomous person-following. This paper highlights and organizes these into a categorical study; to further assist readers in navigating the large body of literature, it is presented in an ordered and organized fashion in Table 6. This section analyzes a number of prominent person-following systems and provides a comparative discussion in qualitative terms. A summary of this qualitative analysis is given in Table 7.

4.1. Detection and tracking performance

An important consideration in designing a perception module is the desired level of detection accuracy and tracking performance given the operating constraints. This impacts the choices of sensors and on-board computational hardware as well. For instance, person-following UGVs can accommodate several sensors, e.g., combinations of cameras, sonars, laser scanners, and RGBD cameras. Therefore, it is generally good practice to adopt sensor fusion schemes (Nikdel et al., 2018; Susperregi et al., 2013; Wang et al., 2018a) to ensure accurate feature-based detection and tracking at a fast rate. If only a single exteroceptive sensor is (or can be) used, more sophisticated techniques, such as deep visual models or online learning-based models, are required to ensure reliable perception (Chen et al., 2017b; Wang et al., 2018b); these models are computationally demanding and typically require single-board supercomputers (NVIDIA, 2014) for real-time inference. However, if there are constraints on power, the use of UWB or RFID tags (Germa et al., 2010; Laneurit et al., 2016) is ideal for designing effective low-power solutions.

The constraints on power consumption and resource utilization are more important considerations for person-following AUVs and UAVs. Hence, using domain-specific prior knowledge, such as modeling divers' swimming patterns by AUVs (Islam and Sattar, 2017) and perspective filtering by UAVs (De Smedt et al., 2015), can facilitate the design of efficient trackers. Nevertheless, on-board supercomputers (NVIDIA, 2014) or edge devices (Google, 2018) can be used to run deep visual trackers in real time (Islam et al., 2018a; Skydio, 2018). Moreover, paired connectivity with the companion human, e.g., paired GPS receivers by UAVs (Staaker, 2016) or acoustic links by ASVs (Miskovic et al., 2015), can provide reliable and fast tracking performances at a low power budget.

The methodological details of these perception modules are discussed in Section 3.1. A qualitative comparison of them is provided in Table 7. The comparison also includes two other important features, i.e., whether online learning is used (Gupta et al., 2017; Park and Kuipers, 2013) and whether person re-identification or recovery is considered (Chen et al., 2017a; Doisy et al., 2012). Additionally, for diver-following systems, invariance to divers' appearance, motion, and wearables is taken into account for comparison. While interpreting this table, it is to be noted that several check-mark (✓) symbols in the first comparison (i.e., for detection and tracking) represent the quality of a proposed solution on a scale of one to three, where three indicates state-of-the-art performance. In all other columns of Table 7, the check-mark (✓) and cross (×) symbols independently represent *yes*, and *no*, respectively, for their corresponding comparisons.

4.2. Optimal planning and control

A few application-specific requirements, particularly the degree of autonomy and the presence of dynamic agents or obstacles in the operating environment directly influence the design choices in planning and control modules of a person-following robot. For instance, in predominately

Table 6. An ordered collection of the person-following systems discussed in this paper; they are mentioned in reverse chronological order and grouped according to their primary focus (perception, planning and control, or interaction). Diamonds (◇) indicate that the corresponding techniques are not specifically about person-following robots, yet are applicable or relevant.

	Perception	Planning and control	Interaction
Ground (2010–2018)	Chen et al. (2017a); Chi et al. (2018); Gupta et al. (2017); Jiang et al. (2018); Popov et al. (2018); Wang et al. (2018a); Chen et al. (2017e)◇; Chen et al. (2017b); Do Hoang et al. (2017); Wang et al. (2017); Cao et al. (2017)◇; Guevara et al. (2016); Koide and Miura (2016); Faria et al. (2015)◇; Babaians et al. (2015); Cai and Matsumaru (2014); Eisenbach et al. (2015); Ilias et al. (2014); Isobe et al. (2014); Leigh et al. (2015); Pairo et al. (2013); Pradhan et al. (2013)◇; Alvarez-Santos et al. (2012); Awai et al. (2013); Basso et al. (2013); Cao and Hashimoto (2013); Chung et al. (2012); Gascueña and Fernández-Caballero (2011); Munaro et al. (2013); Satake et al. (2012, 2013); Susperregi et al. (2013); Yoon et al. (2013); Dollár et al. (2010)◇; Brookshire (2010); Germa et al. (2010)	Chen (2018); Huskić et al. (2017); Nikdel et al. (2018); Pierre (2018); Wang et al. (2018b); Chen et al. (2017d)◇; Masuzawa et al. (2017); Mi et al. (2016); Peng et al. (2016); Cosgun et al. (2016)◇; Sung and Chung (2016); DiGiacomcantonio and Gebreyes (2014); Park and Kuipers (2013); Tarokh and Shenoy (2014); Pradhan et al. (2013)◇; Doisy et al. (2012); Jung and Yi (2012); Morioka et al. (2012); Tarokh and Merloti (2010); Yamaoka et al. (2010)	Honig et al. (2018); Ahn et al. (2018)◇; Fleishman et al. (2018); Pourmehr et al. (2017)◇; Alves-Oliveira and Paiva (2016); Shanee et al. (2016); Triebel et al. (2016); Dewantara and Miura (2016)◇; Thomason et al. (2015)◇; Alvarez-Santos et al. (2014); Hu et al. (2014); The-5elementsrobotics (2014); Cifuentes et al. (2014)◇; Moon et al. (2014)◇; Cosgun et al. (2013); Doisy et al. (2013); Ferrer et al. (2013)◇; Granata and Bidaud (2012); Marge et al. (2011)
Ground (2000–2009)	Hu et al. (2007, 2009); Dollár et al. (2009)◇; Bajracharya et al. (2009)◇; Calisi et al. (2007); Chen and Birchfield (2007); Germa et al. (2009); Handa et al. (2008); Itoh et al. (2006); Kobilarov et al. (2006); Liem et al. (2008); Satake and Miura (2009); Shaker et al. (2008); Takemura et al. (2007); Yoshimi et al. (2006); Zender et al. (2007); Zhu et al. (2006)◇; Kwolek (2004); Kwon et al. (2005); Sedighi et al. (2004)◇; Hirai and Mizoguchi (2003)	Luo et al. (2009); Satake and Miura (2009); Müller et al. (2008)◇; Chivilo et al. (2004); Hoeller et al. (2007); Tarokh and Ferrari (2003)	Yamaoka et al. (2008)◇ ; Gockley et al. (2007); Kuno et al. (2007); Syrdal et al. (2007)◇; Hüttenrauch et al. (2006)◇; Yoshikawa et al. (2006)◇; Kulykukin et al. (2004); Matsumaru et al. (2005)Sakagami et al. (2002)◇
Ground (*–1999)	Stauffer and Grimson (1999)◇ ; Darrell et al. (1998); Schlegel et al. (1998); Papageorgiou et al. (1998)◇; Yamane et al. (1998)◇; Wren et al. (1997)◇; Azarbayejani and Pentland (1996)◇	Sidenbladh et al. (1999); Stentz (1994)◇; Espiau et al. (1992)◇	Piaggio et al. (1998); Burgard et al. (1999)◇; Burgard et al. (1998)◇
Underwater (2010–2018)	Islam and Sattar (2017); Islam et al. (2018a); DeMarco et al. (2013); Gemba et al. (2014); Hari et al. (2015); Mandic et al. (2016); Miskovic et al. (2015)	Islam et al. (2018a); Zadeh et al. (2016)◇; Shkurti et al. (2017)◇; Meger et al. (2014)◇Janabi-Sharifi et al. (2011)◇	Chiarella et al. (2018); Gomez Chavez et al. (2018); Islam et al. (2018b,c); Fulton et al. (2019)◇; Chiarella et al. (2015)◇; Stilinović et al. (2015)
Underwater (*–2009)	Lennartsson et al. (2009); Sattar and Dudek (2009b); Sattar and Dudek (2006)◇	Sattar and Dudek (2009a); Corke et al. (2007)◇; Rosenblatt et al. (2002)◇	Sattar and Dudek (2009b); Xu et al. (2008)◇; Dudek et al. (2007)◇
Aerial	Mueller et al. (2016); Skydio (2018); Vasconcelos and Vasconcelos (2016); Chakrabarty et al. (2016)◇; Barták and Vykovský; (2015)◇; De Smedt et al. (2015); Graether and Mueller (2012); Higuchi et al. (2011); Naseer et al. (2013); Pestana et al. (2014); Portmann et al. (2014); Kumar et al. (2011)◇; Teuliere et al. (2011)◇; Gaszczak et al. (2011)◇	Skydio (2018); Staaker (2016); Huh et al. (2013)◇; Lugo and Zell (2014)◇; Gong et al. (2011)◇; Tomic et al. (2012)◇; Teuliere et al. (2011)◇; Kim et al. (2008)◇	Bruce et al. (2016); Cauchard et al. (2015); Monajjemi et al. (2016); Nagy and Vaughan (2017); Vasconcelos and Vasconcelos (2016); Gioioso et al. (2014)◇; Lichtenstern et al. (2012)◇; Tisdale et al. (2009)◇; Mezouar and Chaumette (2002)◇

Table 7. Qualitative comparisons of a number of prominent person-following systems reported over the last decade (2009–2019). They are compared based on a subset of these items: (i) detection & tracking: qualitative performance; (ii) online: whether an online learning (person re-identification or recovery) module is used; (iii) optimal planning or control: optimality of underlying planning and control modules; (iv) obstacle avoidance: presence of *obstacle avoidance* feature (for UGVs or UAVs); (v) explicit (implicit): availability of some forms of explicit (implicit) interaction; (vi) interactive: availability of interactive user interfaces (for UGVs or UAVs); (vii) multi-H: applicability of the system for multiple-human-following; (viii) outdoors: applicability outdoors (for UGVs or UAVs); (ix) socially aware: availability of socially compliant planning or interaction modules (for UGVs); (x) crowded places: applicability in crowded or populated area (for UGVs or UAVs); (xi) invariance to: whether the tracking performance is invariant to divers'*appearance, motion, and wearables* (of AUV or ASVs); (xii) coastal waters: applicability in coastal and shallow waters (for AUVs or ASVs); (xiii) visibility: applicability in conditions of *poor or no* visibility (for AUVs or ASVs); (xiv) GPS-denied: applicability in GPS-denied environments (for UAVs).

(a) Person-following systems for UGVs.

	Perception, planning, & control			Interaction			Multi-H support & general applicability			
	Detection & tracking	Online (re-identification)	Optimal planning or control	Obstacle avoidance	Explicit (implicit)	Interactive	Multi-H	Outdoors	Socially aware	Crowded places
Wang et al. (2018a)	✓✓✓	× (×)	× / ×	×	× (×)	×	×	✓	×	✓
Nikdel et al. (2018)	✓✓	× (×)	✓/✓	✓	✓(✓)	×	×	×	×	×
Chen (2018)	✓✓✓	× (×)	× /✓	×	× (×)	×	×	×	×	✓
Chen et al. (2017b)	✓✓✓	✓(×)	× /✓	×	× (×)	×	×	✓	×	×
Gupta et al. (2017)	✓✓✓	✓(✓)	× /✓	×	× (✓)	×	×	✓	×	✓
Chen et al. (2017a)	✓✓✓	✓(✓)	× / ×	×	× (✓)	×	×	✓	×	✓
Huskić et al. (2017)	✓✓	× (×)	✓/✓	✓	× (✓)	×	×	✓	×	✓
Koide and Miura (2016)	✓✓✓	✓(✓)	× / ×	×	× (✓)	×	×	✓	×	✓
Triebel et al. (2016)	✓✓✓	✓(×)	✓/ ×	✓	✓(✓)	✓	✓	×	✓	✓
Sung and Chung (2016)	✓✓✓	× (×)	× / ×	×	× (✓)	×	×	×	×	✓
Leigh et al. (2015)	✓✓	× (×)	× /✓	×	× (✓)	×	✓	✓	×	×
Eisenbach et al. (2015)	✓✓✓	× (✓)	× / ×	✓	× (✓)	×	×	×	×	✓
Hu et al. (2014)	✓✓	× (×)	✓/✓	✓	× (✓)	×	×	×	✓	×
Munaro et al. (2013)	✓✓✓	✓(✓)	× / ×	✓	× (✓)	×	✓	×	×	✓
Park and Kuipers (2013)	✓✓	✓(×)	✓/ ×	✓	× (✓)	×	×	✓	✓	✓
Cosgun et al. (2013)	✓✓	× (×)	✓/ ×	✓	✓(✓)	✓	×	×	✓	×
Chung et al. (2012)	✓✓✓	× (✓)	× / ×	×	× (✓)	×	×	×	×	✓
Doisy et al. (2012)	✓✓	× (✓)	✓/✓	✓	× (✓)	×	×	×	×	×
Granata and Bidaud (2012)	✓✓	× (×)	✓/✓	✓	✓(✓)	✓	×	×	✓	✓
Germa et al. (2010)	✓✓✓	× (✓)	× /✓	×	× (✓)	×	×	×	×	✓

(*continued*)

Table 7. Continued

(b) Person-following systems for AUVs (systems for ASVs are marked with an asterisk (*))

	Perception, planning, & control			Interaction	Multi-H support & feasibility		
	Detection & tracking	Invariance to: <appearance, motion, wearables> | online (re-identification)	Optimal planning or control	Human-to-robot (robot-to-human)	Multi-H	Visibility: poor/no	Coastal waters
Islam et al. (2018a)	✓✓✓	<✓, ✓, ✓> | × (×)	✓/✓	✓(×)	✓	✓/×	✓
Islam et al. (2018c)	✓✓✓	<✓, ✓, ✓> | × (×)	✓/✓	×(×)	×	✓/×	✓
Islam and Sattar (2017)	✓✓	<✓, ×, ✓> | × (×)	×/×	×(×)	×	×/×	✓
Mandic et al. (2016)	✓✓✓	<✓, ✓, ✓> | × (×)	×/✓	×(×)	×	✓/✓	×
Hari et al. (2015)	✓✓	<✓, ✓, ✓> | × (×)	×/×	×(×)	×	✓/✓	✓
Miskovic et al. (2015)*	✓✓✓	<✓, ✓, ✓> | × (✓)	✓/✓	×(✓)	✓	✓/✓	×
Gemba et al. (2014)	✓✓	<✓, ✓, ✓> | × (×)	×/×	×(×)	×	✓/✓	✓
DeMarco et al. (2013)	✓✓	<✓, ✓, ✓> | × (×)	×/×	×(×)	×	✓/✓	✓
Sattar and Dudek (2009b)	✓✓	<✓, ×, ×> | × (×)	×/✓	×(×)	×	×/×	✓

(c) Person-following systems for UAVs (commercially available UAVs are marked with an asterisk (*))

	Perception, planning, & control			Interaction			Multi-H support & general applicability			
	Detection & tracking	Online (re-identification)	Optimal planning or control	Obstacle avoidance	Explicit (implicit)	Interactive	Multi-H	Outdoors	GPS-denied	Crowded places
Skydio (2018)*	✓✓✓	×(✓)	✓/✓	✓	✓(×)	✓	✓	✓	×	✓
Vasconcelos and Vasconcelos (2016)	✓✓	×(✓)	×/×	×	✓(×)	×	×	✓	×	×
Mueller et al. (2016)	✓✓✓	×(✓)	×/✓	×	×(×)	✓	×	✓	×	✓
De Smedt et al. (2015)	✓✓✓	×(×)	×/✓	×	×(×)	×	×	✓	×	×
Portmann et al. (2014)	✓✓✓	×(×)	×/×	×	×(×)	×	×	✓	×	✓
Pestana et al. (2014)	✓✓	×(✓)	×/✓	×	×(×)	×	×	✓	✓	✓
Naseer et al. (2013)	✓✓	×(×)	×/✓	×	✓(×)	×	×	×	✓	✓
Staaker (2016)*	✓✓✓	×(✓)	✓/✓	×	✓(×)	×	×	✓	×	✓
Higuchi et al. (2011)	✓✓	×(×)	×/✓	×	✓(×)	×	×	✓	×	✓

ASV: autonomous surface vehicle; AUV: autonomous underwater vehicle; GPS: Global Positioning System; UAV: unmanned aerial vehicle; UGV: unmanned ground vehicle.

static settings, robots can rely on their human companions for collision-free navigation, i.e., plan to maintain a constant distance while assuming that there will be no interfering agents along the way. This approach, often with additional features for obstacle avoidance, is feasible in underwater scenarios (Islam et al., 2018c), and adopted in many ground applications (Koide and Miura, 2016; Sung and Chung, 2016) of person-following. However, as discussed in Section 3.2, optimal planning with consideration of dynamic obstacles, motion, and interaction from other humans, norms of social or public places, etc., is essential for robots operating in crowded area (Granata and Bidaud, 2012; Park and Kuipers, 2013), social settings (Cosgun et al., 2013; Triebel et al., 2016), and challenging outdoor scenarios (Mueller et al., 2016; Staaker, 2016).

Conversely, complex motion planning requires dense knowledge about the environment, which impacts the choice and modality of sensors. For instance, UGVs operating in known indoor environments can take advantage of a global map (Nikdel et al., 2018) in order to accurately plan to navigate while avoiding obstacles (Triebel et al., 2016). Even when a global map is not available, 3D sensing capabilities (e.g., a camera with sonar, LRF, or infrared sensors, or several cameras) are needed to obtain localized 3D information about the world, which can be used for SLAM-based navigation (Huskić et al., 2017; Skydio, 2018). Furthermore, based on application-specific requirements, the rules of social norms and desired implicit behaviors of the robot must be modeled as prior knowledge and eventually incorporated into planning and control modules. These aspects are also considered in the qualitative comparison given in Table 7.

4.3. Interactivity and general feasibility

A number of important design choices depend on the desired level of interactivity between a robot and its companion human (Cosgun et al., 2013; Granata and Bidaud, 2012). This influences the choice of sensors or peripheral devices (interactive screen, voice interface, paired application, etc.), and the design of important perception and planning modules (hand-gesture recognition, action recognition, planning for implicit interaction, etc.). Additionally, some aspects, such as whether multiple-human support (i.e., following as a group) or social awareness is needed and the choice between following ahead or behind, are essential considerations while designing a person-following system. These interactivity requirements need to be formulated by thorough user experiments for practical applications (Gockley et al., 2007; Triebel et al., 2016).

Several features pertaining to the interactivity and general feasibility of person-following robots are considered for qualitative comparison in Table 7. These aspects, relevant design issues based on various use cases, and the corresponding state-of-the-art solutions for ground, underwater, and aerial scenarios are elaborately discussed in this paper. As is evident from these discussions, the vast majority of the literature on person-following robots addresses various research problems in ground scenarios. It is safe to say that the current state-of-the-art systems provide very good solutions to these problems. However, the social and behavioral aspects of these systems require more attention from researchers. In addition, better and smarter methodologies are required to address the unique challenges of underwater and aerial scenarios. These aspects, and other important research directions are highlighted in the following section.

5. Prospective research directions

The following subsections discuss a number of active research areas and open problems that are naturally challenging and are potentially useful in person-following applications.

5.1. Following a team

Many underwater missions involve a team of several divers working together (Figure 18). Following the team as a whole is operationally more efficient in general. Similar scenarios arise when filming a social or sports event using UAVs. The perception problem can be easily solved by a simple extension (i.e., by allowing the detection of several humans); however, motion planning and control modules are not straightforward. Moreover, the rules for spatial conduct and interaction need to be identified and quantified. Tracking a team of independently moving objects is a challenging problem in general (Shu et al., 2012); it gets even more challenging in a 3D environment while dealing with real-time constraints. Despite the challenges, it is potentially invaluable in numerous applications of person-following robots (Shkurti et al., 2012; Wellbots, 2015).

5.2. Following as a team (convoying)

Multi-robot human-led convoys are useful in cooperative estimation problems (Rekleitis et al., 2001). A simple approach to this problem is to assign leader–follower pairs; that is, one of the robots is assigned to follow the person, and every other robot is individually assigned another robot as its leader. Each robot follows its leader and together they move as a team. Another approach is to let the robots communicate with each other and cooperatively plan their motions. The underlying planning pipeline is similar to that of a multi-robot convoying problem, which is particularly challenging in underwater and aerial scenarios (Minaeian et al., 2016; Shkurti et al., 2017). Moreover, this can be further generalized into the problem of *following a group of people by a team of autonomous robots* in a cooperative setting. However, a complex cooperative planning pipeline is required to achieve optimal positioning and motion trajectories for each robot, which is an open problem as well.

5.3. Following behind or ahead?

There are scenarios where it is ideal to have the robot stay ahead of the person while following. Hands-free shopping-cart robots, for instance, should stay ahead of the human, not behind (Kuno et al., 2007; Nikdel et al., 2018). Another prime example is of person-following UAVs that record sports activities; they should be able to move around and take snapshots from different directions to get the best perspective (Skydio, 2018). Therefore, traditional systems and methodologies for *following from behind* are not very useful in these applications.

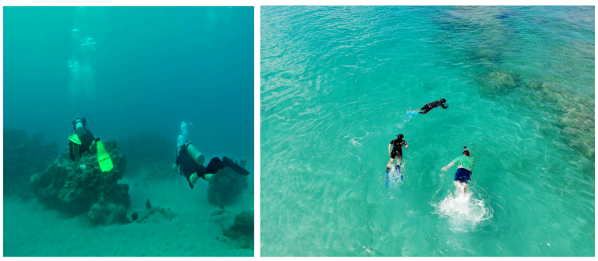
In recent years, researchers have begun to explore the particularities of different scenarios (Figure 19) where the robot should be in front or at the side of the person while following (Ferrer et al., 2013; Hu et al., 2014; Nagy and Vaughan, 2017). These scenarios impart more operational challenges since the robot needs to predict the motion trajectory of the person, and needs some way to recover from a wrong prediction or action. Knowledge of motion history and gaze behaviors of the person, and prior knowledge about the environment or destination can be utilized to model such anticipative behaviors. The person can help the robot make decisions in critical situations as well (using hand gestures or voice commands). Nevertheless, these aspects demand more research attention and experimental evaluations in real-world settings.

5.4. Learning to follow from demonstration

End-to-end learning of autonomous robot behaviors from demonstration is an interesting ongoing research topic. Researchers have reported exciting results in the domains of 2D robot navigation in cluttered environments (Pfeiffer et al., 2017), simple autonomous driving (Codevilla et al., 2018), imitating driving styles (Kuderer et al., 2015), etc. These results indicate that the end-to-end learning models, particularly the idea of *learning from demonstration* can be very effective for person-following robots. Further research attention is required to explore other end-to-end (deep) learning-based models as well because they have the potential to significantly simplify autonomous person-following. There are a few research efforts already in this regard in simulation environments (Dewantara and Miura, 2016; Pierre, 2018); however, more extensive research and real-world experiments are necessary.

5.5. Human–robot communication

A generic communication paradigm for human–robot dialog (Thomason et al., 2015) can be very useful in practice for person-following applications. Several human-to-robot communication paradigms using speech, markers, and hand gestures are discussed in this paper. There are not many research studies on how a robot can initiate communication and maintain a proper dialog with the human, particularly in applications where interactive user interfaces are not feasible (Fulton et al., 2019). Furthermore, effective and efficient risk assessment in human–robot dialog (Robinette et al., 2016; Sattar and Dudek, 2011) is another potential research problem in this domain.



(a) Underwater scenario. **(b)** Aerial scenario.

Fig. 18. Views from robots' cameras while following teams of people.

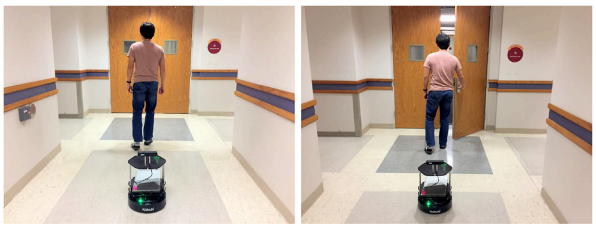
(a) UAV filming an athlete from various viewpoints (Skydio, 2018). **(b)** UGV leading person through hallway.

Fig. 19. Scenarios where a robot is not following its companion from behind.

5.6. Enabling social and spatial awareness

Various forms of implicit human–robot interaction, particularly the preferred spatial and motion behaviors for person-following robots were discussed in the previous section. Robots that are deployed in a social setting should be *aware* of these aspects and the social norms in general (Granata and Bidaud, 2012; Honig et al., 2018; Kim and Mutlu, 2014).

A particular instance of anticipative robot behavior is illustrated in Figure 20. Here, the robot anticipates the door-opening action (Zender et al., 2007), increases the distance from the person by slowing down, and waits instead of moving forward. Many other anticipated behaviors, such as moving slowly while entering cross-paths, waiting at a side when the person is interacting with other people, etc., are important features of a social robot. These are difficult to quantify and implement in general (Chen et al., 2017d; Kruse et al., 2013); extensive experiments and further user studies are required to model these social norms for person-following robots.



(a) UGV following person while staying behind.

(b) UGV standing clear of door opening.

Fig. 20. Desired robot behavior: notice that the UGV is giving extra space to the person to open the door.

5.7. Long-term interaction and support

Another social aspect of the person-following UGV is long-term interaction with a human companions. This has numerous potential applications in health care; for instance, Coninx et al. (2016) showed that long-term child–robot interaction was useful for learning and therapeutic purposes; Chen et al. (2017c) and Kidd and Breazeal (2008) proved that long-term interaction with a robot helped people in physical exercises. These, among many other studies, show that robots can help more by learning about the general behaviors and routine activities of their human companions. Thorough analyses and user studies are needed to discover the feasibilities and utilities of long-term interactions for other person-following applications.

5.8. Specific person-following

Following a *specific* person is generally more useful than following *any* person, specially in a multi-human setting (Satake and Miura, 2009) and in social or crowded environments. Moreover, the ability to follow a specific person is an essential feature for UGVs that accompany older people and people with disabilities (Ilias et al., 2014; Liem et al., 2008). It is straightforward to achieve this in some applications, with the use of an additional human face or body-pose recognition module (Cao and Hashimoto, 2013; Yoshimi et al., 2006). However, scenarios such as following a person in crowded surrounding (Germa et al., 2009) or avoiding an impeding person (Hoeller et al., 2007) are rather challenging. Furthermore, lack of person-specific features while viewing a diver from behind (different divers may wear similar suits), make it a harder problem for underwater robots (Xia and Sattar, 2019). Detecting a specific person from a distant UAV is also challenging for similar reasons.

5.9. Person re-identification

Several mechanisms for person *recovery* or *re-identification* used by existing person-following systems are mentioned in Section 3.2.5. They mostly use feature-based template-matching (Do Hoang et al., 2017; Gupta et al., 2017; Koide and Miura, 2016) techniques; trajectory replication-based techniques (Chen et al., 2017a) are also used for re-identification when the target person transiently disappears from the robot's view and appears again. A number of recently proposed appearance-based deep models (Ahmed et al., 2015; Li et al., 2014) have significantly improved the state-of-the-art performance for person re-identification on standard datasets. Despite the potentials, these models are yet to be used in person-following systems. Investigating the applicability of these person re-identification models for specific person-following in human-dominated social settings is an interesting and potentially rewarding research direction.

5.10. Surveillance and rescue support

Features such as person re-identification and adversarial person-following are useful for autonomous human surveillance using UAVs (Portmann et al., 2014). Additionally, in human rescue missions, a team of UAVs is invaluable in adversarial conditions (Doherty and Rudol, 2007). These are critical applications and there is always room for improvements.

5.11. Embedded parallel computing solutions

As mentioned earlier, deep learning-based models provide robust solutions to most of the perception problems involved in person-following scenarios. One practical limitation of these models is that they are often computationally expensive and require parallel computing platforms. Therefore, faster mobile supercomputers and embedded parallel computing solutions (Google, 2018; NVIDIA, 2014) will be immensely useful in person-following applications. The recent success of the person-following UAV named Skydio R1 (Skydio, 2018) is a practical example. However, the high power consumption of these on-board computers is still a major concern; for instance, the flight time for a Skydio R1 is only about 16 min. In addition to computational capacity and power consumption, many other aspects of mobile supercomputers, such as durability and cooling mechanisms, require further technological improvements. Future advancements in ultra-low-power computer vision (TinyVision) and machine learning (TinyML) techniques and platforms (Warden and Situnayake, 2019) might play an important role in this regard.

5.12. Addressing privacy and safety concerns

There have been an increasing number of concerns across cyberspace about the privacy and safety issues of autonomous robots, particularly UAVs operating in social and public environments (UCTV, 2013). A recent study (Hitlin, 2017) has found that about 54% of the US population thinks that drones and autonomous UAVs should not be allowed to fly near people's homes. This is because the use

of drones undermines people's ability to assess context and measure trust. While person-following UAVs are mostly used for recreational purposes in public areas and often crowded places, these concerns need to be addressed using technological and educational solutions (Finn and Wright, 2012; Wang et al., 2016) to ensure transparency and trust.

6. Conclusions

Person-following by autonomous robots has numerous important applications in industry. In addition, the usage of person-following robots in social settings and for entertainment purposes has flourished over the last decade. Researchers have approached various aspects of the *autonomous person-following* problem from different perspectives and contributed to the development of a vast body of literature. This paper makes an effort to present a comprehensive overview of this large body of literature in a categorical fashion. First, design issues and operational challenges for person-following robots in ground, underwater, and aerial scenarios are presented. Then state-of-the-art methods for perception, planning, control, and interaction of various person-following systems are elaborately discussed.

In addition, several operational considerations for applying these methods, underlying assumptions, and their feasibility in different use cases is analyzed and compared in qualitative terms. Finally, a number of open problems and potential applications are highlighted for future research; improved solutions to these problems will significantly strengthen the literature and bridge the gap between research and practice.

Acknowledgements

We would like to thank all the reviewers for their insightful comments and suggestions, which immensely enriched this paper. We are also thankful to Arindam Banerjee (professor, University of Minnesota), Richard T Vaughan (associate professor, Simon Fraser University), and Florian Shkurti (assistant professor, University of Toronto) for their valuable insights and directions. Lastly, we acknowledge the contributions of our colleagues, namely Michael Fulton, Marcus Oh, Cameron Fabbri, Julian Lagman, Marc Ho, Youya Xia, Elliott Imhoff, Peigen Luo, and Yuyang Xiao for their assistance in collecting data, annotating images, and preparing media files.

ORCID iD

Md Jahidul Islam https://orcid.org/0000-0001-7211-2675

References

Ahmed E, Jones M and Marks TK (2015) An improved deep learning architecture for person re-identification. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Boston, MA, USA, 7–12 June 2015, pp. 3908–3916. Piscataway, NJ: IEEE.

Ahn J, Kim M, Kim S, et al. (2018) Formation-based tracking method for human following robot. In: *15th international conference on ubiquitous robots (UR)*, Honolulu, HI, USA, 26–30 June 2018, pp. 24–28. Piscataway, NJ: IEEE.

Alvarez-Santos V, Iglesias R, Pardo XM, et al. (2014) Gesture-based interaction with voice feedback for a tour-guide robot. *Journal of Visual Communication and Image Representation* 25(2): 499–509.

Alvarez-Santos V, Pardo XM, Iglesias R, et al. (2012) Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot. *Robotics and Autonomous Systems* 60(8): 1021–1036.

Alves-Oliveira P and Paiva A. (2016) A study on trust in a robotic suitcase. In: Agah A, Cabibihan JJ, Howard A, et al. (eds.) *Social Robotics. ICSR 2016 (Lecture Notes in Computer Science*, vol. 9979). Cham: Springer, p. 179.

Awai M, Shimizu T, Kaneko T, et al. (2013) HOG-based person following and autonomous returning using generated map by mobile robot equipped with camera and laser range finder. In: Lee S, Cho H, Yoon KJ, et al. (eds) *Intelligent Autonomous Systems 12 (Advances in Intelligent Systems and Computing*, vol. 194). Berlin: Springer, pp. 51–60.

Azarbayejani A and Pentland A. (1996) Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features. In: *13th international conference on pattern recognition*, Vienna, Austria, 25–29 August 1996, vol. 3, pp. 627–632. Piscataway, NJ: IEEE.

Babaians E, Korghond NK, Ahmadi A, et al. (2015) Skeleton and visual tracking fusion for human following task of service robots. In: *RSI international conference on robotics and mechatronics (ICROM)*, Tehran, Iran, 7–9 October 2015, pp. 761–766. Piscataway, NJ: IEEE.

Bajracharya M, Moghaddam B, Howard A, et al. (2009) A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle. *International Journal of Robotics Research (IJRR)* 28(11–12): 1466–1485.

Balan AO, Sigal L and Black MJ (2005) A quantitative evaluation of video-based 3D person tracking. In: *IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance*, Beijing, China, 15–16 October 2005, pp. 349–356. Piscataway, NJ: IEEE.

Barták R and Vykovský A. (2015) Any object tracking and following by a flying drone. In: *Mexican international conference on artificial intelligence (MICAI)*, Cuernavaca, Mexico, 25–31 October 2015, pp. 35–41. Piscataway, NJ: IEEE.

Basso F, Munaro M, Michieletto S, et al. (2013) Fast and robust multi-people tracking from RGB-D data for a mobile robot. *Intelligent Autonomous Systems* 12: 265–276.

Brookshire J. (2010) Person following using histograms of oriented gradients. *International Journal of Social Robotics* 2(2): 137–146.

Bruce J, Monajjemi V, Wawerla J, et al. (2016) Tiny people finder: Long-range outdoor HRI by periodicity detection. In: *13th conference on computer and robot vision (CRV)*, Victoria, Canada, 1–3 June 2016, pp. 216–221. Piscataway, NJ: IEEE.

Burgard W, Cremers AB, Fox D, et al. (1998) The interactive museum tour-guide robot. In: *AAAI '98/IAAI '98 proceedings of the fifteenth national/tenth conference on artificial intelligence/innovative applications of artificial intelligence*, Madison, Wisconsin, USA, 26–30 July 1998, pp. 11–18. Menlo Park, CA: American Association for Artificial Intelligence.

Burgard W, Cremers AB, Fox D, et al. (1999) Experiences with an interactive museum tour-guide robot. *Artificial Intelligence* 114(1–2): 3–55.

Cacace J, Finzi A and Lippiello V. (2016) Multimodal interaction with multiple co-located drones in search and rescue missions. *arXiv* arXiv:1605.07316.

Cai J and Matsumaru T. (2014) Human detecting and following mobile robot using a laser range sensor. *Journal of Robotics and Mechatronics* 26(6): 718–734.

Calisi D, Iocchi L and Leone R. (2007) Person following through appearance models and stereo vision using a mobile robot. In: *VISAPP (workshop on robot vision)*, Barcelona, Spain, 8–11 March 2007, pp. 46–56. Setubal: Institute for Systems and Technologies of Information, Control and Communication.

Cao M and Hashimoto H (2013) Specific person recognition and tracking of mobile robot with Kinect 3D sensor. In: *39th annual conference of the IEEE industrial electronics society, (IECON)*, Vienna, Austria, 10–13 November 2013, pp. 8323–8328. Piscataway, NJ: IEEE.

Cao Z, Simon T, Wei SE, et al. (2017) Realtime multi-person 2D pose estimation using part affinity fields. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017, pp. 7291–7299. Piscataway, NJ: IEEE.

Cauchard JR, Zhai KY and Landay JA. (2015) Drone & me: an exploration into natural human–drone interaction. In: *ACM international joint conference on pervasive and ubiquitous computing*, Osaka, Japan, âĂŤ 7–11 September 2015, pp. 361–365. New York, NY: ACM

Chakrabarty A, Morris R, Bouyssounouse X, et al. (2016) Autonomous indoor object tracking with the Parrot AR. drone. In: *International conference on unmanned aircraft systems (ICUAS)*, Arlington, VA, USA, 7–10 June 2016, pp. 25–30. Piscataway, NJ: IEEE.

Chen BX, Sahdev R and Tsotsos JK. (2017a) Integrating stereo vision with a CNN tracker for a person-following robot. In: Liu M, Chen H and Vincze M. (eds.) *Computer Vision Systems. ICVS 2017 (Lecture Notes in Computer Science*, vol. 10528). Cham: Springer, Cham, pp. 300–313.

Chen BX, Sahdev R and Tsotsos JK. (2017b) Person following robot using selected online Ada-boosting with stereo camera. In: *14th conference on computer and robot vision (CRV)*, Edmonton, Canada, 16–19 May 2017, pp. 48–55. Piscataway, NJ: IEEE.

Chen E. (2018) "FOLO": A vision-based human-following robot. In: *3rd international conference on automation, mechanical control and computational engineering (AMCCE)* (ed. Cai M. and Zheng G.), 12–13 May 2018, Dalian, China. Paris: Atlantis Press.

Chen Y, Garcia-Vergara S and Howard AM. (2017c) Effect of feedback from a socially interactive humanoid robot on reaching kinematics in children with and without cerebral palsy: A pilot study. *Developmental Neurorehabilitation* 21(8): 490–496.

Chen YF, Everett M, Liu M, et al. (2017d) Socially aware motion planning with deep reinforcement learning. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vancouver, Canada, 24–28 September 2017, pp. 1343–1350. Piscataway, NJ: IEEE.

Chen YF, Liu M, Everett M, et al. (2017e) Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In: *IEEE international conference on robotics and automation (ICRA)*, Singapore, 29 May–3 June 2017, pp. 285–292. Piscataway, NJ: IEEE.

Chen Z and Birchfield ST. (2007) Person following with a mobile robot using binocular feature-based tracking. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, San Diego, CA, USA, 29 October–2 November 2007, pp. 815–820. Piscataway, NJ: IEEE.

Chi W, Wang J and Meng MQH (2018) A gait recognition method for human following in service robots. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48(9): 1429–1440.

Chiarella D, Bibuli M, Bruzzone G, et al. (2015) Gesture-based language for diver-robot underwater interaction. In: *OCEANS 2015—Genova*, Genoa, Italy, 18–21 May 2015, pp. 1–9. Piscataway, NJ: IEEE.

Chiarella D, Bibuli M, Bruzzone G, et al. (2018) A novel gesture-based language for underwater human–robot interaction. *Journal of Marine Science and Engineering* 6(3): 91.

Chivilo G, Mezzaro F, Sgorbissa A, et al. (2004) Follow-the-leader behaviour through optical flow minimization. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Sendai, Japan, 28 September–2 October 2004, vol. 4, pp. 3182–3187. Piscataway, NJ: IEEE.

Chung W, Kim H, Yoo Y, et al. (2012) The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder. *IEEE Transactions on Industrial Electronics* 59(8): 3156–3166.

Cifuentes CA, Frizera A, Carelli R, et al. (2014) Human–robot interaction based on wearable IMU sensor and laser range finder. *Robotics and Autonomous Systems* 62(10): 1425–1439.

Codevilla F, Miiller M, López A, et al. (2018) End-to-end driving via conditional imitation learning. In: *IEEE international conference on robotics and automation (ICRA)*, Brisbane, Australia, 21–25 May 2018, pp. 1–9. Piscataway, NJ: IEEE.

Coninx A, Baxter P, Oleari E, et al. (2016) Towards long-term social child–robot interaction: Using multi-activity switching to engage young users. *Journal of Human–Robot Interaction* 5(1): 32–67.

Corke P, Detweiler C, Dunbabin M, et al. (2007) Experiments with underwater robot localization and tracking. In: *IEEE international conference on robotics and automation (ICRA)*, Rome, Italy, 10–14 April 2007, pp. 4556–4561. Piscataway, NJ: IEEE.

Cosgun A, Florencio DA and Christensen HI (2013) Autonomous person following for telepresence robots. In: *IEEE international conference on robotics and automation (ICRA)*, Karlsruhe, Germany, 6–10 May 2013, pp. 4335–4342. Piscataway, NJ: IEEE.

Cosgun A, Sisbot EA and Christensen HI. (2016) Anticipatory robot path planning in human environments. In: *25th international symposium on robot and human interactive communication (RO-MAN)*, New York, NY, USA, 26–31 August 2016, pp. 562–569. Piscataway, NJ: IEEE.

Cu G, Ang AG, Ballesteros AR, et al. (2013) Human following robot using Kinect sensor. In: *Research congress*, Manila, Philippines, 7–9 March 2013, pp. 1–7. Manila, Philippines: DLSUPH.

Darrell T, Gordon G, Harville M, et al. (1998) Integrated person tracking using stereo, color, and pattern detection. In: *IEEE computer society conference on computer vision and pattern recognition (CVPR)*, Santa Barbara, CA, USA, 25 June 1998, pp. 601–608. Piscataway, NJ: IEEE.

De Smedt F, Hulens D and Goedemé T (2015) On-board real-time tracking of pedestrians on a UAV. In: *IEEE conference on*

computer vision and pattern recognition (CVPR) workshops, Boston, MA, USA, 7–12 June 2015, pp. 1–8. Piscataway, NJ: IEEE.

De Wit CC, Siciliano B and Bastin G. (2012) *Theory of Robot Control*. London: Springer Science & Business Media.

DeMarco KJ, West ME and Howard AM (2013) Sonar-based detection, and tracking of a diver for underwater human–robot interaction scenarios. In: *IEEE international conference on systems, man, and cybernetics (SMC)*, Manchester, UK, 13–16 October 2013, pp. 2378–2383. Piscataway, NJ: IEEE.

Dewantara BSB and Miura J (2016) Generation of a socially aware behavior of a guide robot using reinforcement learning. In: *International electronics symposium (IES)*, Denpasar, Indonesia, 29–30 September 2016, pp. 105–110. Piscataway, NJ: IEEE.

DiGiacomcantonio M and Gebreyes Y. (2014) *Self-propelled luggage*. Patent application US2014/0107868-A1, USA.

DJI (2015) Inspire 2.0: Power beyond imagination. Available at: https://www.dji.com/inspire-2/. (accessed: 30 January 2019).

DJI (2016) Mavic Pro: Wherever you go. Available at: https://www.dji.com/mavic. (accessed: 30 January 2019).

Do Hoang M, Yun SS and Choi JS (2017) The reliable recovery mechanism for person-following robot in case of missing target. In: *International conference on ubiquitous robots and ambient intelligence (URAI)*, Jeju, South Korea, 28 June–1 July 2017, pp. 800–803. Piscataway, NJ: IEEE.

Doherty P and Rudol P. (2007) A UAV search and rescue scenario with human body detection and geolocalization. In: Orgun MA and Thornton J (eds.) *AI 2007: Advances in Artificial Intelligence. AI 2007 (Lecture Notes in Computer Science*, vol. 4830). Berlin: Springer, pp. 1–13.

Doisy G, Jevtić A and Bodiroža S. (2013) Spatially unconstrained, gesture-based human–robot interaction. In: *8th ACM/IEEE international conference on human–robot interaction*, Tokyo, Japan, 3–6 March 2013, pp. 117–118. Piscataway, NJ: IEEE.

Doisy G, Jevtic A, Lucet E, et al. (2012) Adaptive person-following algorithm based on depth images and mapping. In: *IROS workshop on robot motion planning: online, reactive, and in real-time*, Vilamoura, Portugal, 7–12 October 2012. Piscataway, NJ: IEEE.

Dollár P, Belongie SJ and Perona P (2010) The fastest pedestrian detector in the West. In: *British machine vision conference*, Aberystwyth, UK, 31 August–3 September 2010, vol. 2, p. 7. Durham: BMVA Press.

Dollár P, Tu Z, Perona P, et al. (2009) Integral channel features. In: *British machine vision conference*, 7–10 September 2009, London, UK, pp. 91.1–91.11. Durham: BMVA Press.

Dudek G, Sattar J and Xu A (2007) A visual language for robot control, and programming: A human-interface study. In: *IEEE international conference on robotics and automation (ICRA)*, Rome, Italy, 10–14 April 2007, pp. 2507–2513. Piscataway, NJ: IEEE.

Eisenbach M, Vorndran A, Sorge S, et al. (2015) User recognition for guiding and following people with a mobile robot in a clinical environment. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Hamburg, Germany, 28 September–2 October 2015, pp. 3600–3607. Piscataway, NJ: IEEE.

Espiau B, Chaumette F and Rives P. (1992) A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation* 8(3): 313–326.

Ess A, Leibe B, Schindler K, et al. (2008) A mobile vision system for robust multi-person tracking. In: *IEEE conference on computer vision and pattern recognition (CVPR)*, Anchorage, AK, USA, 23–28 June 2008, pp. 1–8. Piscataway, NJ: IEEE.

Faria DR, Vieira M, Premebida C, et al. (2015) Probabilistic human daily activity recognition towards robot-assisted living. In: *24th IEEE international symposium on robot and human interactive communication (RO-MAN)*, Kobe, Japan, 31 August–4 September 2015, pp. 582–587. Piscataway, NJ: IEEE.

Ferrer G, Garrell A and Sanfeliu A. (2013) Robot companion: A social-force based approach with human awareness-navigation in crowded environments. In: *2013 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Tokyo, Japan, 3–7 November 2013, pp. 1688–1694. Piscataway, NJ: IEEE.

Finn RL and Wright D. (2012) Unmanned aircraft systems: Surveillance, ethics and privacy in civil applications. *Computer Law & Security Review* 28(2): 184–194.

Fleishman V, Honig S, Oron-Gilad T, et al. (2018) Proxemic preferences when being followed by a robot. Israeli Ministry of Science and Technology, Report: 3–12060.

Fritsch J, Kleinehagenbrock M, Lang S, et al. (2004) Audiovisual person tracking with a mobile robot. In: *International conference on intelligent autonomous systems* (eds. Groen F, Amato N, Bonarini A, et al.), Amsterdam, Netherlands, 10–12 March 2004, pp. 898–906. Amsterdam: IOS Press.

Fulton M, Edge C and Sattar J (2019) Robot communication via motion: Closing the underwater human–robot interaction loop. *2019 international conference on robotics and automation (ICRA)*, Montreal, Canada, 20–24 May 2019. Piscataway, NJ: IEEE.

Gascueña JM and Fernández-Caballero A. (2011) Agent-oriented modeling and development of a person-following mobile robot. *Expert Systems with Applications* 38(4): 4280–4290.

Gaszczak A, Breckon TP and Han J. (2011) Real-time people and vehicle detection from UAV imagery. In: *SPIE 7878, intelligent robots and computer vision XXVIII: algorithms and techniques* (eds. Röning J, Casasent DP and Hall EL), San Francisco, CA, USA, 24–25 January 2011, p. 78780B. Birmingham, WA: International Society for Optics and Photonics.

Gemba KL, Nosal EM and Reed TR. (2014) Partial dereverberation used to characterize open circuit SCUBA diver signatures. *Journal of the Acoustical Society of America* 136(2): 623–633.

Germa T, Lerasle F, Ouadah N, et al. (2009) Vision and RFID-based person tracking in crowds from a mobile robot. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, St. Louis, MO, USA, 10–15 October 2009, pp. 5591–5596. Piscataway, NJ: IEEE.

Germa T, Lerasle F, Ouadah N, et al. (2010) Vision and RFID data fusion for tracking people in crowds by a mobile robot. *Computer Vision and Image Understanding* 114(6): 641–651.

Ghosh A, Alboul L, Penders J, et al. (2014) Following a robot using a haptic interface without visual feedback, In: *7th international conference on advances in computer–human interactions, ACHI 2014*, Barcelona, Spain, 23–27 March 2014. Wilmington, DE: IARIA XPS Press.

Gioioso G, Franchi A, Salvietti G, et al. (2014) The flying hand: A formation of UAVs for cooperative aerial tele-manipulation. In: *IEEE international conference on robotics and automation (ICRA)*, Hong Kong, China, 31 May–7 June 2014, pp. 4335–4341. Piscataway, NJ: IEEE.

Gockley R, Forlizzi J and Simmons R (2007) Natural person-following behavior for social robots. In: *ACM/IEEE international conference on human–robot interaction*, Arlington, VA, USA, 10–12 March 2007, pp. 17–24. New York, NY: ACM.

Goldhoorn A, Garrell A, Alquézar R, et al. (2014) Continuous real-time POMCP to find-and-follow people by a humanoid service robot. In: *IEEE-RAS international conference on humanoid robots*, Madrid, Spain, 18–20 November 2014, pp. 741–747. Piscataway, NJ: IEEE.

Gomez Chavez A, Ranieri A, Chiarella D, et al. (2018) CADDY underwater stereo-vision dataset for human–robot interaction (HRI) in the context of diver activities. *Journal of Marine Science and Engineering* 7(1): 16.

Gong H, Sim J, Likhachev M, et al. (2011) Multi-hypothesis motion planning for visual object tracking. In: *IEEE international conference on computer vision (ICCV)*, Barcelona, Spain, 6–13 November 2011, pp. 619–626. Piscataway, NJ: IEEE.

González D, Pérez J, Milanés V, et al. (2015) A review of motion planning techniques for automated vehicles. *IEEE Transactions on Intelligent Transportation Systems* 17(4): 1135–1145.

Google (2018) Coral USB accelerator. Available at: coral.withgoogle.com/products/accelerator/. (accessed: 20 June 2019).

Graether E and Mueller F. (2012) Joggobot: A flying robot as jogging companion. In: *CHI extended abstracts on human factors in computing systems*, Austin, TX, USA, 5–10 May 2012, pp. 1063–1066. New York, NY: ACM.

Granata C and Bidaud P (2012) A framework for the design of person following behaviors for social mobile robots. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vilamoura, Portugal, 7–12 October 2012, pp. 4652–4659. Piscataway, NJ: IEEE.

Guevara AE, Hoak A, Bernal JT, et al. (2016) Vision-based self-contained target following robot using Bayesian data fusion. In: Bebis G, Boyle R, Parvin B, et al. (eds.) *Advances in Visual Computing. ISVC 2016 (Lecture Notes in Computer Science*, vol. 10072). Cham: Springer, pp. 846–857.

Gupta M, Kumar S, Behera L, et al. (2017) A novel vision-based tracking algorithm for a human-following mobile robot. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47(7): 1415–1427.

Handa A, Sivaswamy J, Krishna KM, et al. (2008) Person following with a mobile robot using a modified optical flow. In: Marques L, de Almeida A, Tokhi MO, et al. *Advances in Mobile Robotics*. Singapore: World Scientific, pp. 1154–1160.

Hari VN, Chitre M, Too YM, et al. (2015) Robust passive diver detection in shallow ocean. In: *OCEANS 2015—Genova*, Genoa, Italy, 18–21 May 2015, pp. 1–6. Piscataway, NJ: IEEE.

Higuchi K, Shimada T and Rekimoto J (2011) Flying sports assistant: External visual imagery representation for sports training. In: *Augmented human international conference*, Tokyo, Japan, 13 March 2011, no. 7. New York, NY: ACM.

Hirai N and Mizoguchi H (2003) Visual tracking of human back and shoulder for person following robot. In: *IEEE/ASME international conference on advanced intelligent mechatronics*, Kobe, Japan, 20–24 July 2003, pp. 527–532. Piscataway, NJ: IEEE.

Hitlin P. (2017) 8% of Americans say they own a drone, while more than half have seen one in operation. Available at: http://www.pewresearch.org/staff/paul-hitlin/. (accessed: 2 February 2019).

Hoeller F, Schulz D, Moors M, et al. (2007) Accompanying persons with a mobile robot using motion prediction and probabilistic roadmaps. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, San Diego, CA, USA, 29 October–2 November 2007, pp. 1260–1265. Piscataway, NJ: IEEE.

Honig SS, Oron-Gilad T, Zaichyk H, et al. (2018) Towards socially aware person-following robots. *IEEE Transactions on Cognitive and Developmental Systems* 10(4): 936–954.

Hu C, Ma X and Dai X (2007) A robust person tracking, and following approach for mobile robot. In: *International conference on mechatronics and automation*, Harbin, China, 5–8 August 2007, pp. 3571–3576. Piscataway, NJ: IEEE.

Hu CH, Ma XD and Dai XZ (2009) Reliable person following approach for mobile robot in indoor environment. In: *International conference on machine learning and cybernetics*, Hebei, China, 12–15 July 2009, vol. 3, pp. 1815–1821. Piscataway, NJ: IEEE.

Hu JS, Wang JJ and Ho D. (2014) Design of sensing system and anticipative behavior for human following of mobile robots. *IEEE Transactions on Industrial Electronics* 61(4): 1916–1927.

Huh S, Shim DH and Kim J (2013) Integrated navigation system using camera and gimbaled laser scanner for indoor, and outdoor autonomous flight of UAVs. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Tokyo, Japan, 3–7 November 2013, pp. 3158–3163. Piscataway, NJ: IEEE.

Huskić G, Buck S, González LAI, et al. (2017) Outdoor person following at higher speeds using a skid-steered mobile robot. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vancouver, Canada, 24–28 September 2017, pp. 3433–3438. Piscataway, NJ: IEEE.

Hüttenrauch H, Eklundh KS, Green A, et al. (2006) Investigating spatial relationships in human–robot interaction. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Beijing, China, 9–15 October 2006, pp. 5052–5059. Piscataway, NJ: IEEE.

Ilias B, Shukor SA, Yaacob S, et al. (2014) A nurse following robot with high speed Kinect sensor. *ARPN Journal of Engineering and Applied Sciences* 9(12): 2454–2459.

Inoue H, Tachikawa T and Inaba M (1992) Robot vision system with a correlation chip for real-time tracking optical flow, and depth map generation. In: *IEEE international conference on robotics and automation (ICRA)*, Nice, France, 12–14 May 1992, pp. 1621–1626. Piscataway, NJ: IEEE.

Iribe M, Matsuda H, Aizawa H, et al. (2011) Study on a practical robotic follower to support daily life—Mobile robot development for home oxygen therapy patients with the hyper tether. *Journal of Robotics and Mechatronics* 23(2): 316–323.

Islam MJ and Sattar J (2017) Mixed-domain biological motion tracking for underwater human–robot interaction. In: *IEEE international conference on robotics and automation (ICRA)*, Singapore, 29 May–3 June 2017, pp. 4457–4464. Piscataway, NJ: IEEE.

Islam MJ, Fulton M and Sattar J. (2018a) Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection. *IEEE Robotics and Automation Letters* 4(1): 113–120.

Islam MJ, Ho M and Sattar J (2018b) Dynamic reconfiguration of mission parameters in underwater human–robot collaboration. In: *2018 IEEE international conference on robotics and*

automation (ICRA), Brisbane, Australia, 21–25 May 2018. Piscataway, NJ: IEEE.

Islam MJ, Ho M and Sattar J. (2018c) Understanding human motion and gestures for underwater human–robot collaboration. *Journal of Field Robotics* 36(5): 851–873.

Isobe Y, Masuyama G and Umeda K (2014) Human following with a mobile robot based on combination of disparity and color images. In: *Europe–Asia congress on mecatronics (MECATRONICS)*, Tokyo, Japan, 27–29 November 2014, pp. 84–88. Piscataway, NJ: IEEE.

Itoh K, Kikuchi T, Takemura H, et al. (2006) Development of a person following mobile robot in complicated background by using distance and color information. In: *IEEE annual conference on industrial electronics*, Paris, France, 6–10 November 2006, pp. 3839–3844. Piscataway, NJ: IEEE.

Janabi-Sharifi F, Deng L and Wilson W. (2011) Comparison of basic visual servoing methods. *IEEE/ASME Transactions on Mechatronics* 16(5): 967–983.

Jiang S, Yao W, Hong Z, et al. (2018) A classification-lock tracking strategy allowing a person-following robot to operate in a complicated indoor environment. *Sensors* 18(11): 3903.

Julier SJ and Uhlmann JK (1997) New extension of the Kalman filter to nonlinear systems. In: *AeroSense*, Orlando, FL, USA, 28 July 1997, pp. 182–193. Birmingham, WA: International Society for Optics and Photonics.

Jung EJ and Yi BJ (2012) Control algorithms for a mobile robot tracking a human in front. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vilamoura, Portugal, 7–12 October 2012, pp. 2411–2416. Piscataway, NJ: IEEE.

Kalman R. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* 82(1): 35–45.

Kanda T, Shiomi M, Miyashita Z, et al. (2009) An affective guide robot in a shopping mall. In: *ACM/IEEE international conference on human robot interaction*, La Jolla, CA, USA, 9–13 March 2009, pp. 173–180. New York, NY: ACM.

Kidd CD and Breazeal C (2008) Robots at home: Understanding long-term human–robot interaction. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Nice, France, 22–26 September 2008, pp. 3230–3235. Piscataway, NJ: IEEE.

Kim Y and Mutlu B. (2014) How social distance shapes human–robot interaction. *International Journal of Human-Computer Studies* 72(12): 783–795.

Kim Y, Gu DW and Postlethwaite I. (2008) Real-time path planning with limited information for autonomous unmanned air vehicles. *Automatica* 44(3): 696–712.

Kobilarov M, Sukhatme G, Hyams J, et al. (2006) People tracking and following with mobile robot using an omnidirectional camera and a laser. In: *IEEE international conference on robotics and automation (ICRA)*, Orlando, FL, USA, 15–19 May 2006, pp. 557–562. Piscataway, NJ: IEEE.

Koide K and Miura J. (2016) Identification of a specific person using color, height, and gait features for a person following robot. *Robotics and Autonomous Systems* 84: 76–87.

Kruse T, Pandey AK, Alami R, et al. (2013) Human-aware robot navigation: A survey. *Robotics and Autonomous Systems* 61(12): 1726–1743.

Kuderer M, Gulati S and Burgard W (2015) Learning driving styles for autonomous vehicles from demonstration. In: *IEEE international conference on robotics and automation (ICRA)*, Seattle, WA, USA, 26–30 May, 2015, pp. 2641–2646. Piscataway, NJ: IEEE.

Kulykukin V, Gharpure C and DeGraw N (2004) Human–robot interaction in a robotic guide for the visually impaired. In: *AAAI spring symposium*, Palo Alto, CA, USA, 22–24 March 2004, pp. 158–164. Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Kumar KS, Kavitha G, Subramanian R, et al. (2011) Visual and thermal image fusion for UAV based target tracking. In: Ionescu CM. (ed.) *MATLAB—A Ubiquitous Tool for the Practical Engineer*. Rijeka: InTech.

Kuno Y, Sadazuka K, Kawashima M, et al. (2007) Museum guide robot based on sociological interaction analysis. In: *SIGCHI conference on human factors in computing systems*, San Jose, CA, USA, 28 April–3 May 2007, pp. 1191–1194. New York, NY: ACM.

Kwolek B. (2004) Person following and mobile camera localization using particle filters. In: *4th international workshop on robot motion and control*, Puszczykowo, Poland, 20 June 2004, pp. 265–270. Piscataway, NJ: IEEE.

Kwon H, Yoon Y, Park JB, et al. (2005) Person tracking with a mobile robot using two uncalibrated independently moving cameras. In: *IEEE international conference on robotics and automation*, Barcelona, Spain, 18–22 April 2005, pp. 2877–2883. Piscataway, NJ: IEEE.

Laneurit J, Chapuis R and Debain C. (2016) TRACKBOD, an accurate, robust and low cost system for mobile robot person following. In: *5th international conference on machine control and guidance (MCG)*, Vichy, France, 5–6 October 2016.

Leigh A, Pineau J, Olmedo N, et al. (2015) Person tracking and following with 2D laser scanners. In: *IEEE international conference on robotics and automation (ICRA)*, Seattle, WA, USA, 26–30 May 2015, pp. 726–733. Piscataway, NJ: IEEE.

Lennartsson R, Dalberg E, Fristedt T, et al. (2009) Electric detection of divers in harbor environments. In: *OCEANS*, Biloxi, MS, USA, 26–29 October 2009, pp. 1–8. Piscataway, NJ: IEEE.

Li W, Zhao R, Xiao T, et al. (2014) Deepreid: Deep filter pairing neural network for person re-identification. In: *27th IEEE conference on computer vision and pattern recognition (CVPR)*, Columbus, OH, USA, 24–27 June 2014, pp. 152–159. Washington, DC: IEEE Computer Society.

Lichtenstern M, Frassl M, Perun B, et al. (2012) A prototyping environment for interaction between a human and a robotic multi-agent system. In: *ACM/IEEE international conference on human–robot interaction*, Boston, MA, USA, 5–8 March 2012, pp. 185–186. New York, NY: ACM.

Liem M, Visser A and Groen F (2008) A hybrid algorithm for tracking, and following people using a robotic dog. In: *3rd ACM/IEEE international conference on human–robot interaction (HRI)*, Amsterdam, The Netherlands, 12–15 March 2008, pp. 185–192. New York, NY: ACM.

Liu W, Anguelov D, Erhan D, et al. (2016) SSD: Single shot multibox detector. In: Leibe B, Matas J, Sebe N, et al. (eds.) *Computer Vision—ECCV 2016 (Lecture Notes in Computer Science*, vol. 9905). Cham: Springer, pp. 21–37.

Lugo JJ and Zell A. (2014) Framework for autonomous on-board navigation with the AR drone. *Journal of Intelligent & Robotic Systems* 73(1-4): 401–412.

Luo RC, Chang NW, Lin SC, et al. (2009) Human tracking and following using sensor fusion approach for mobile assistive companion robot. In: *Annual conference of IEEE on industrial*

electronics, Porto, Portugal, 3–5 November 2009, pp. 2235–2240. Piscataway, NJ: IEEE.

Ma X, Hu C, Dai X, et al. (2008) Sensor Integration for Person Tracking and Following with Mobile Robot. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Porto, Portugal, 3–5 November 2009, pp. 3254–3259. Piscataway, NJ: IEEE.

Mandic F, Rendulic I, Miskovic N, et al. (2016) Underwater object tracking using sonar and USBL measurements. *Journal of Sensors* 2016: 8070286.

Marge M, Powers A, Brookshire J, et al. (2011) Comparing heads-up, hands-free operation of ground robots to teleoperation. *Robotics: Science and systems VII* (eds. Durrant-Whyte H, Roy N and Abbeel P), Los Angeles, CA, USA, 27–30 June 2011, Cambridge, MA: MIT Press.

Masuzawa H, Miura J and Oishi S (2017) Development of a mobile robot for harvest support in greenhouse horticulture—Person following and mapping. In: *IEEE/SICE international symposium on system integration (SII)*, Taipei, Taiwan, 11–14 December 2017, pp. 541–546. Piscataway, NJ: IEEE.

Matsumaru T, Iwase K, Akiyama K, et al. (2005) Mobile robot with eyeball expression as the preliminary-announcement and display of the robot's following motion. *Autonomous Robots* 18(2): 231–246.

Meger D, Shkurti F, Poza DC, et al. (2014) 3D trajectory synthesis and control for a legged swimming robot. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Chicago, IL, USA, 14–18 September 2014, pp. 2257–2264. Piscataway, NJ: IEEE.

Mezouar Y and Chaumette F. (2002) Path planning for robust image-based control. *IEEE Transactions on Robotics and Automation* 18(4): 534–549.

Mi W, Wang X, Ren P, et al. (2016) A system for an anticipative front human following robot. In: *International conference on artificial intelligence and robotics and the international conference on automation, control and robotics engineering*, Kitakyushu, Japan, 13–15 July 2016, no. 4. New York, NY: ACM.

Minaeian S, Liu J and Son Y. (2016) Vision-based target detection and localization via a team of cooperative UAV and UGVs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 46(7): 1005–1016.

Miskovic N, Nad D and Rendulic I. (2015) Tracking divers: An autonomous marine surface vehicle to increase diver safety. *IEEE Robotics and Automation Magazine* 22(3): 72–84.

Monajjemi M, Mohaimenianpour S and Vaughan R. (2016) UAV, come to me: End-to-end, multi-scale situated HRI with an uninstrumented human and a distant UAV. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Daejeon, South Korea, 9–14 October 2016, pp. 4410–4417. Piscataway, NJ: IEEE.

Moon A, Troniak DM, Gleeson B, et al. (2014) Meet me where I'm gazing: How shared attention gaze affects human–robot handover timing. In: *ACM/IEEE international conference on human–robot interaction*, Bielefeld, Germany, 3–6 March 2014, pp. 334–341. New York, NY: ACM.

Morales M, Tapia L, Pearce R, et al. (2004) A machine learning approach for feature-sensitive motion planning. In: Erdmann M, Overmars M, Hsu D, et al. (eds.) *Algorithmic Foundations of Robotics VI (Springer Tracts in Advanced Robotics*, vol. 17). Berlin: Springer, pp. 361–376.

Morioka K, Oinaga Y and Nakamura Y. (2012) Control of human-following robot based on cooperative positioning with an intelligent space. *Electronics and Communications in Japan* 95(1): 20–30.

Mueller M, Sharma G, Smith N, et al. (2016) Persistent aerial tracking system for UAVs. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Daejeon, South Korea, 9–14 October 2016, pp. 1562–1569. Piscataway, NJ: IEEE.

Müller J, Stachniss C, Arras KO, et al. (2008) Socially inspired motion planning for mobile robots in populated environments. In: *International conference on cognitive systems*, Karlsruhe, Germany, 2–4 April 2008.

Munaro M, Basso F, Michieletto S, et al. (2013) A software architecture for RGB-D people tracking based on ROS framework for a mobile robot. In: Lee S, Yoon KJ and Lee J. (eds) *Frontiers of Intelligent Autonomous Systems (Studies in Computational Intelligence*, vol. 466). Berlin: Springer, pp. 53–68.

Nagy G and Vaughan R. (2017) Flying face engagement: Aligning a UAV to directly face a moving uninstrumented user. In: *IEEE international conference on intelligent robots and systems (IROS'17)*, Vancouver, Canada, 24–28 September 2017. Piscataway, NJ: IEEE.

Naseer T, Sturm J and Cremers D. (2013) FollowMe: Person following and gesture recognition with a quadrocopter. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Tokyo, Japan, 3–7 November 2013, pp. 624–630. Piscataway, NJ: IEEE.

Nikdel P, Shrestha R and Vaughan R (2018) The hands-free push-cart: Autonomous following in front by predicting user trajectory around obstacles. In: *IEEE international conference on robotics and automation (ICRA)*, Brisbane, Australia, 21–25 May 2018. Piscataway, NJ: IEEE.

Nishimura S, Takemura H and Mizoguchi H (2007) Development of attachable modules for robotizing daily items—person following shopping cart robot. In: *IEEE international conference on robotics and biomimetics*, Sanya, China, 15–18 December 2007, pp. 1506–1511. Piscataway, NJ: IEEE.

NVIDIA (2014) NVIDIA developer blog. Available at: https://devblogs.nvidia.com/ (accessed 2 February 2019).

Pairo W, Ruiz-del Solar J, Verschae R, et al. (2013) Person following by mobile robots: Analysis of visual and range tracking methods and technologies. In: Behnke S, Veloso M, Visser A, et al. (eds) *RoboCup 2013: Robot World Cup XVII (Lecture Notes in Computer Science*, vol. 8371). Berlin: Springer, pp. 231–243.

Papageorgiou CP, Oren M and Poggio T (1998) A general framework for object detection. In: *International conference on computer vision*, Bombay, India, 7 January 1998, pp. 555–562. Piscataway, NJ: IEEE.

Park CH and Howard AM. (2010) Towards real-time haptic exploration using a mobile robot as mediator. In: *IEEE haptics symposium*, Waltham, MA, USA, 25–26 March 2010, pp. 289–292. Piscataway, NJ: IEEE.

Park JJ and Kuipers B. (2013) Autonomous person pacing and following with model predictive equilibrium point control. In: *IEEE international conference on robotics and automation (ICRA)*, Karlsruhe, Germany, 6–10 May 2013, pp. 1060–1067. Piscataway, NJ: IEEE.

Parrot S. (2012) AR.Drone 2.0. Available at: https://www.parrot.com/us/drones/. (accessed 30 January 2019).

Peng W, Wang J and Chen W. (2016) Tracking control of human-following robot with sonar sensors. In: Chen W, Hosoda K, Menegatti E, et al. (eds.) *Intelligent Autonomous Systems 14.*

IAS 2016 (Advances in Intelligent Systems and Computing, vol. 531). Cham: Springer, pp. 301–313.

Pestana J, Sanchez-Lopez JL, Saripalli S, et al. (2014) Computer vision based general object following for GPS-denied multirotor unmanned vehicles. In: *American control conference (ACC)*, Portland, OR, USA, 4–6 June 2014, pp. 1886–1891. Piscataway, NJ: IEEE.

Pfeiffer M, Schaeuble M, Nieto J, et al. (2017) From perception to decision: A data-driven approach to end-to-end motion planning for autonomous ground robots. In: *IEEE international conference on robotics and automation (ICRA)*, Singapore, 29 May–3 June 2017, pp. 1527–1533. Piscataway, NJ: IEEE.

Piaggio M, Fornaro R, Piombo A, et al. (1998) An optical-flow person following behaviour. In: *Intelligent control (ISIC) held jointly with IEEE international symposium on computational intelligence in robotics and automation (CIRA), intelligent systems and semiotics (ISAS)*, Gaithersburg, MD, USA, 17 September 1998, pp. 301–306. Piscataway, NJ: IEEE.

Piaggio-Fast-Forward (2017) Gita. Available at: http://piaggiofastforward.com/ (accessed: 2 February 2019).

Pierre JM (2018) End-to-end deep learning for robotic following. In: *2nd international conference on mechatronics systems and control engineering*, Amsterdam, Netherlands, 21–23 February 2018, pp. 77–85. New York, NY: ACM.

Pieska S, Luimula M, Jauhiainen J, et al. (2013) Social service robots in wellness and restaurant applications. *Journal of Communication and Computer* 10(1): 116–123.

Popov VL, Ahmed SA, Shakev NG, et al. (2018) Detection and following of moving targets by an indoor mobile robot using Microsoft Kinect and 2D lidar data. In: *International conference on control, automation, robotics and vision (ICARCV)*, Singapore, 18–21 November 2018, pp. 280–285. Piscataway, NJ: IEEE.

Portmann J, Lyne S, Chli M, et al. (2014) People detection and tracking from aerial thermal views. In: *IEEE international conference on robotics and automation (ICRA)*, Hong Kong, China, 31 May–7 June 2014, pp. 1794–1800. Piscataway, NJ: IEEE.

Pounds P, Mahony R and Corke P. (2010) Modelling and control of a large quadrotor robot. *Control Engineering Practice* 18(7): 691–699.

Pourmehr S, Thomas J, Bruce J, et al. (2017) Robust sensor fusion for finding HRI partners in a crowd. In: *IEEE international conference on robotics and automation (ICRA)*, Singapore, 29 May–3 June 2017, pp. 3272–3278. Piscataway, NJ: IEEE.

Pradhan N, Burg T, Birchfield S, et al. (2013) Indoor navigation for mobile robots using predictive fields. In: *American control conference (ACC)*, Washington, DC, USA, 17–19 June 2013, pp. 3237–3241. Piscataway, NJ: IEEE.

Pucci D, Marchetti L and Morin P (2013) Nonlinear control of unicycle-like robots for person following. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Tokyo, Japan, 3–7 November 2013, pp. 3406–3411. Piscataway, NJ: IEEE.

Rekleitis I, Dudek G and Milios E. (2001) Multi-robot collaboration for robust exploration. *Annals of Mathematics and Artificial Intelligence* 31(1–4): 7–40.

Robinette P, Wagner AR and Howard A. (2016) Investigating human–robot trust in emergency scenarios: Methodological lessons learned. In: Mittu R, Sofge D, Wagner A, et al. (eds.) *Robust Intelligence and Trust in Autonomous Systems*. Boston, MA: Springer, pp. 143–166.

Rosenblatt J, Williams S and Durrant-Whyte H. (2002) A behavior-based architecture for autonomous underwater exploration. *Information Sciences* 145(1): 69–87.

Rudol P and Doherty P (2008) Human body detection and geolocalization for UAV search, and rescue missions using color, and thermal imagery. In: *IEEE aerospace conference*, Big Sky, MT, USA, 1–8 March 2008, pp. 1–8. Piscataway, NJ: IEEE.

Sakagami Y, Watanabe R, Aoyama C, et al. (2002) The intelligent ASIMO: System overview and integration. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Lausanne, Switzerland, 30 September–4 October 2002, vol. 3, pp. 2478–2483. Piscataway, NJ: IEEE.

Satake J and Miura J (2009) Robust stereo-based person detection and tracking for a person following robot. In: *ICRA workshop on people detection and tracking*, Kobe, Japan, 12–17 May 2009, pp. 1–10. Piscataway, NJ: IEEE.

Satake J, Chiba M and Miura J (2012) A SIFT-based person identification using a distance-dependent appearance model for a person following robot. In: *IEEE international conference on robotics and biomimetics (ROBIO)*, Guangzhou, China, 11–14 December 2012, pp. 962–967. Piscataway, NJ: IEEE.

Satake J, Chiba M and Miura J. (2013) Visual person identification using a distance-dependent appearance model for a person following robot. *International Journal of Automation and Computing* 10(5): 438–446.

Sattar J and Dudek G (2006) On the performance of color tracking algorithms for underwater robots under varying lighting and visibility. *IEEE international conference on robotics and automation. ICRA 2006*, Orlando, FL, USA, 15–19 May 2006, pp. 3550–3555. Piscataway, NJ: IEEE.

Sattar J and Dudek G (2009a) Robust servo-control for underwater robots using banks of visual filters. In: *IEEE international conference on robotics and automation (ICRA)*, Kobe, Japan, 12–17 May 2009, pp. 3583–3588. Piscataway, NJ: IEEE.

Sattar J and Dudek G. (2009b) Underwater human–robot interaction via biological motion identification. In: *Robotics: science and systems V.* Seattle, WA, USA, 28 June–1 July 2009.Cambridge, MA: MIT Press.

Sattar J and Dudek G (2011) Towards quantitative modeling of task confirmations in human–robot dialog. In: *IEEE international conference on robotics and automation (ICRA)*, Shanghai, China, 9–13 May 2011, pp. 1957–1963. Piscataway, NJ: IEEE.

Sattar J, Bourque E, Giguere P, et al. (2007) Fourier tags: Smoothly degradable fiducial markers for use in human–robot interaction. In: *Canadian conference on computer and robot vision (CRV)*, Montreal, Canada, 28–30 May 2007, pp. 165–174. Piscataway, NJ: IEEE.

Sattar J, Dudek G, Chiu O, et al. (2008) Enabling autonomous capabilities in underwater robotics. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Nice, France, 22–26 September 2008, pp. 3628–3634. Piscataway, NJ: IEEE.

Schlegel C, Illmann J, Jaberg H, et al. (1998) Vision based person tracking with a mobile robot. In: *British machine vision conference* (eds. Nixon M and Carter J), Southampton, UK, 14–17 September 1998, pp. 15.1–15.10. Durham: BMVA Press.

Sedighi KH, Ashenayi K, Manikas TW, et al. (2004) Autonomous local path planning for a mobile robot using a genetic algorithm. In: *Congress on evolutionary computation*, Portland,

OR, USA, 19–23 June 2004, vol. 2, pp. 1338–1345. Piscataway, NJ: IEEE.

Shaker S, Saade JJ and Asmar D. (2008) Fuzzy inference-based person-following robot. *International Journal of Systems Applications, Engineering & Development* 2(1): 29–34.

Shanee HS, Dror K, Tal OG, et al. (2016) The influence of following angle on performance metrics of a human-following robot. In: *IEEE international symposium on robot and human interactive communication (RO-MAN)*, New York, NY, USA, 26–31 August 2016, pp. 593–598. Piscataway, NJ: IEEE.

Shkurti F, Chang WD, Henderson P, et al. (2017) Underwater multi-robot convoying using visual tracking by detection. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vancouver, Canada, 24–28 September 2017. Piscataway, NJ: IEEE.

Shkurti F, Xu A, Meghjani M, et al. (2012) Multi-domain monitoring of marine environments using a heterogeneous robot team. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Vilamoura, Portugal, 7–12 October 2012, pp. 1747–1753. Piscataway, NJ: IEEE.

Shu G, Dehghan A, Oreifej O, et al. (2012) Part-based multiple-person tracking with partial occlusion handling. In: *IEEE computer society conference on computer vision and pattern recognition (CVPR)*, Providence, RI, USA, 16–21 June 2012, pp. 1815–1821. Piscataway, NJ: IEEE.

Sidenbladh H, Kragic D and Christensen HI (1999) A person following behaviour for a mobile robot. In: *IEEE international conference on robotics and automation (ICRA)*, Detroit, MI, USA, 10–15 May 1999, vol. 1, pp. 670–675. Piscataway, NJ: IEEE.

Skydio. (2018) Skydio R1. Available at: https://www.skydio.com/ (accessed 30 January 2019).

Slabbekoorn H, Bouton N, van Opzeeland I, et al. (2010) A noisy spring: The impact of globally rising underwater sound levels on fish. *Trends in Ecology & Evolution* 25(7): 419–427.

SPi Corp (2015) The SPI infrared cameras. Available at: https://www.x20.org/shop/ (accessed 20 January 2019).

Staaker (2016) All eyes on you. Available at: https://www.staaker.com/ (accessed: 10 February 2019).

Stauffer C and Grimson WEL (1999) Adaptive background mixture models for real-time tracking. In: *IEEE computer society conference on computer vision and pattern recognition (CVPR)*, Fort Collins, CO, USA, 23–25 June 1999, vol. 2, pp. 246–252. Piscataway, NJ: IEEE.

Stentz A (1994) Optimal and efficient path planning for partially-known environments. In: *IEEE international conference on robotics and automation (ICRA)*, San Diego, CA, USA, 8–13 May 1994, pp. 3310–3317. Piscataway, NJ: IEEE.

Stilinović N, Nađ D and Mišković N (2015) AUV for diver assistance, and safety – Design, and implementation. In: *OCEANS 2015—Genova*, Genoa, Italy, 18–21 May 2015, pp. 1–4. Piscataway, NJ: IEEE.

Sung Y and Chung W. (2016) Hierarchical sample-based joint probabilistic data association filter for following human legs using a mobile robot in a cluttered environment. *IEEE Transactions on Human-Machine Systems* 46(3): 340–349.

Susperregi L, Martínez-Otzeta JM, Ansuategui A, et al. (2013) RGB-D, laser and thermal sensor fusion for people following in a mobile robot. *International Journal of Advanced Robotic Systems* 10(6): 271.

Syrdal DS, Koay KL, Walters ML, et al. (2007) A personalized robot companion?—The role of individual differences on spatial preferences in HRI scenarios. In: *IEEE international symposium on robot and human interactive communication*, Jeju, South Korea, 26–29 August 2007, pp. 1143–1148. Piscataway, NJ: IEEE.

Takemura H, Ito K and Mizoguchi H (2007) Person following mobile robot under varying illumination based on distance and color information. In: *IEEE international conference on robotics and biomimetics*, Sanya, China, 15–18 December 2007, pp. 1500–1505. Piscataway, NJ: IEEE.

Tarokh M and Ferrari P. (2003) Case study: Robotic person following using fuzzy control and image segmentation. *Journal of Robotic Systems* 20(9): 557–568.

Tarokh M and Merloti P. (2010) Vision-based robotic person following under light variations and difficult walking maneuvers. *Journal of Field Robotics* 27(4): 387–398.

Tarokh M and Shenoy R. (2014) Vision-based robotic person following in fast walking. In: *IEEE international conference on systems, man and cybernetics*, San Diego, CA, USA, 5–8 October 2014, pp. 3172–3177. Piscataway, NJ: IEEE.

Tasaki R, Kitazaki M, Miura J, et al. (2015) Prototype design of medical round supporting robot 'Terapio'. In: *IEEE international conference on robotics and automation (ICRA)*, Seattle, WA, USA, 26–30 May 2015, pp. 829–834. Piscataway, NJ: IEEE.

Tensorflow (2017) Tensorflow object detection zoo. Available at: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md (accessed 20 February 2019).

Teuliere C, Eck L and Marchand E. (2011) Chasing a moving target from a flying UAV. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, San Francisco, CA, USA, 25–30 September 2011, pp. 4929–4934. Piscataway, NJ: IEEE.

The-5elementsrobotics (2014) Budgee the friendly robot. Available at: http://5elementsrobotics.com/budgee-main/ (accessed 2 February 2019).

Thomason J, Zhang S, Mooney RJ, et al. (2015) Learning to interpret natural language commands through human–robot dialog. In: *24th international joint conference on artificial intelligence*, Buenos Aires, Argentina, 25–31 July 2015, pp. 1923–1929. Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Tisdale J, Kim Z and Hedrick J. (2009) Autonomous UAV path planning and estimation. *IEEE Robotics & Automation Magazine* 16(2): 35–42.

Tomic T, Schmid K, Lutz P, et al. (2012) Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue. *IEEE Robotics & Automation Magazine* 19(3): 46–56.

Triebel R, Arras K, Alami R, et al. (2016) Spencer: A socially aware service robot for passenger guidance and help in busy airports. In: Wettergreen D and Barfoot T. (eds.) *Field and Service Robotics (Springer Tracts in Advanced Robotics*, vol. 113). Cham: Springer, pp. 607–622.

UCTV (2013) Drones and other UAVs: Benefits and risks—exploring ethics. Available at: https://youtu.be/obSzr2NIEEM. (accessed 2 February 2019).

Vasconcelos F and Vasconcelos N. (2016) Person-following UAVS. In: *IEEE winter conference on computer vision (WACV)*, Lake Placid, NY, USA, 7–10 March 2016, pp. 1–9. Piscataway, NJ: IEEE.

Wan EA and Van Der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: *Adaptive systems for signal processing, communications, and control symposium*, Lake Louise, Canada, 4 October 2000, pp. 153–158. Piscataway, NJ: IEEE.

Wang M, Liu Y, Su D, et al. (2018a) Accurate and real-time 3D tracking for the following robots by fusing vision and ultrasonar information. *IEEE/ASME Transactions on Mechatronics* 23(3): 997–1006.

Wang M, Su D, Shi L, et al. (2017) Real-time 3D human tracking for mobile robots with multisensors. *arXiv* arXiv:1703.04877.

Wang X, Zhang L, Wang D, et al. (2018b) Person detection, tracking and following using stereo camera. In: *International conference on graphic and image processing (ICGIP)*, Qingdao, China, 10 April 2018, vol. 10615, p. 106150D. Birmingham, WA: International Society for Optics and Photonics.

Wang Y, Xia H, Yao Y, et al. (2016) Flying eyes and hidden controllers: A qualitative study of people's privacy perceptions of civilian drones in the US. *Privacy Enhancing Technologies* 2016(3): 172–190.

Warden P and Situnayake D. (2019) *TinyML: Machine Learning with TensorFlow on Arduino, and Ultra-Low Power Micro-Controllers*. Sebastopol, CA: O'Reilly Media.

Wellbots (2015) 7 ways to use a drone. Available at: https://www.wellbots.com/7-ways-to-use-a-drone/ (accessed 20 January 2019).

Willow-Garage (2011) Person following by a TurtleBot. Available at: turtlebot.com (accessed 10 March 2019).

Wren CR, Azarbayejani A, Darrell T, et al. (1997) Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7): 780–785.

Wu X, Stuck RE, Rekleitis I, et al. (2015) Towards a framework for human factors in underwater robotics. In: *Human factors and ergonomics society annual meeting*, Los Angeles, CA, USA, 26–30 October 2015, vol. 59, pp. 1115–1119. Thousand Oaks, CA: Sage.

Xia Y and Sattar J (2019) Visual diver recognition for underwater human–robot collaboration. *IEEE international conference on robotics and automation (ICRA)*, Montreal, Canada, 20–24 May 2019. Piscataway, NJ: IEEE.

Xu A, Dudek G and Sattar J (2008) A natural gesture interface for operating robotic systems. In: *IEEE international conference on robotics and automation (ICRA)*, Pasadena, CA, USA, 19–23 May 2008, pp. 3557–3563. Piscataway, NJ: IEEE.

Yamane T, Shirai Y and Miura J. (1998) Person tracking by integrating optical flow and uniform brightness regions. In: *IEEE international conference on robotics and automation (ICRA)*, Leuven, Belgium, 20 May 1998, vol. 4, pp. 3267–3272. Piscataway, NJ: IEEE.

Yamaoka F, Kanda T, Ishiguro H, et al. (2008) How close? Model of proximity control for information-presenting robots. In: *ACM/IEEE international conference on human robot interaction*, Amsterdam, The Netherlands, 12–15 March 2008, pp. 137–144. New York, NY: ACM.

Yamaoka F, Kanda T, Ishiguro H, et al. (2010) A model of proximity control for information-presenting robots. *IEEE Transactions on Robotics* 26(1): 187–195.

Yang L, Qi J, Song D, et al. (2016) Survey of robot 3D path planning algorithms. *Journal of Control Science and Engineering* 2016: 5.

Yoon Y, Yoon H and Kim J. (2013) Depth assisted person following robots. In: *IEEE international symposium on robot and human interactive communication (RO-MAN)*, Gyeongju, South Korea, 26–29 August 2013, pp. 330–331. Piscataway, NJ: IEEE.

Yoon Y, Yun W, Yoon H, et al. (2014) Real-time visual target tracking in RGB-D data for person-following robots. In: *International conference on pattern recognition (ICPR)*, Stockholm, Sweden, 24–28 August 2014, pp. 2227–2232. Piscataway, NJ: IEEE.

Yoshikawa Y, Shinozawa K, Ishiguro H, et al. (2006) Responsive robot gaze to interaction partner. In: *Robotics: science and systems II* (eds. Sukhatme GS, Schaal S, Burgard W, et al.), Philadelphia, PA, 16–19 August 2006. Cambridge, MA, MIT Press.

Yoshimi T, Nishiyama M, Sonoura T, et al. (2006) Development of a person following robot with vision based target detection. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, Beijing, China, 9–15 October 2006, pp. 5286–5291. Piscataway, NJ: IEEE.

Zadeh SM, Yazdani AM, Sammut K, et al. (2016) AUV rendezvous online path planning in a highly cluttered undersea environment using evolutionary algorithms. *arXiv* arXiv:1604.07002.

Zeiler MD and Fergus R. (2014) Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, et al. (eds.) *Computer Vision—ECCV 2014 (Lecture Notes in Computer Science*, vol. 8689). Cham: Springer, pp. 818–833.

Zender H, Jensfelt P and Kruijff GJM (2007) Human and situation-aware people following. In: *IEEE international symposium on robot and human interactive communication*, Jeju, South Korea, 26–29 August 2007, pp. 1131–1136. Piscataway, NJ: IEEE.

Zhu Q, Yeh MC, Cheng KT, et al. (2006) Fast human detection using a cascade of histograms of oriented gradients. In: *IEEE computer society conference on computer vision and pattern recognition (CVPR)*, New York, NY, USA, 17–22 June 2006, vol. 2, pp. 1491–1498. Piscataway, NJ: IEEE.