Article

The International Journal of Robotics Research
1–21

DOI: 10.1177/0278364917707583
journals.sagepub.com/home/ijr

First-person view semi-autonomous teleoperation of cooperative wheeled mobile robots with visuo-haptic feedback

Kent Yee Lui[1,2], Hyunjun Cho[1,3], ChangSu Ha[1] and Dongjun Lee[1]

Abstract

We propose a novel semi-autonomous teleoperation control framework for two nonholonomic wheeled mobile robots, which are cooperatively grasping and transporting a deformable object without any physical grip-enforcing fixture and are teleoperated by a single remote human user via a first-person view camera attached on the over-seeing wheeled mobile robot with some visuo-haptic feedback. First, we apply nonholonomic passive decomposition to split the kinematics of the two wheeled mobile robots into the grasping/first-person view-centering behavior and the teleoperation-related behavior. We reveal that, under a mixed constraint of the system (i.e. physical/nonholonomic no-slip/drift constraint and artificial/holonomic grasping/first-person view-centering constraint), the first-person view camera cannot track an arbitrary human command. We then design the semi-autonomous control architecture, consisting of an autonomous grasping-enforcing and first-person view-centering control and a first-person view camera teleoperation control, which in turns stands upon autonomous human command tracking control. We also design visual and haptic feedback to notify the user of, and also nudge them to reduce, the command-behavior mismatch caused by the mixed constraint, which is difficult for the user to understand via the limited first-person view, we can thus induce user confusion/frustration and consequently performance deterioration. Experiment verification and a human subject study are performed to show the efficacy of our proposed first-person view semi-autonomous teleoperation framework with the visuo-haptic feedback.

Keywords

Behavior decomposition, cooperative grasping and transport, first-person view, nonholonomic wheeled mobile robots, semi-autonomous teleoperation, visuo-haptic feedback

1. Introduction

Even with the recent rapid advancement in artificial intelligent (AI), it is still difficult to attain many practically important robotic tasks that solely relying on fully-autonomous robots. This is particularly so when the environment in which the robots are required to operate, is uncertain, unknown, complex and/or dynamic. For such practical robotic tasks, it is often desired, or imperative, to utilize some teleoperation capacity to impart human intelligence to cope with the tasks that are difficult to achieve fully autonomously (e.g. task priotization/replanning in a dynamically changing environment, recognition of salient features not pre-defined and the decision on what to do with those, the reaction to a contact or impact with their dynamics which is often too complex to model for real-time autonomous reasoning, etc.).

Teleoperation is one of the oldest fields in robotics (Goertz, 1952; Goertz and Bevilacqua, 1952) and has enjoyed strong results, many of which are typically limited to the bilateral teleoperation of master-slave systems consisting of two kinematically similar ground-fixed robotic manipulators (e.g. Anderson and Spong, 1989; Khademian and Hashtrudi-Zaad, 2011; Kim et al., 1992; Lawrence, 1993; Lee and Li, 2005; Lee and Spong, 2006; Pacchierotti et al., 2015). Recently, along with the interests and advancements in robotic mobility and multi-robot systems, some new results have been proposed on the teleoperation of multiple mobile robots (e.g. Franchi et al., 2012a,b; Ha and Lee, 2013; Lee, 2008; Lee and Spong, 2005; Lee et al., 2013). At the same time, the issue of using FPV (first-person view) for teleoperation is emerging, spurred particularly by its recent adoption in drone racing and also by the availability of many affordable HMDs (head-mounted displays).

[1]Department Mechanical and Aerospace Engineering and IAMD, Seoul National University, Republic of Korea
[2]ASML Korea, Republic of Korea
[3]Hanwha Cooperation, Republic of Korea

Corresponding author:
Dongjun Lee, Department of Mechanical & Aerospace Engineering and IAMD, Seoul National University, Seoul, 151–744, Republic of Korea.
Email: djlee@snu.ac.kr

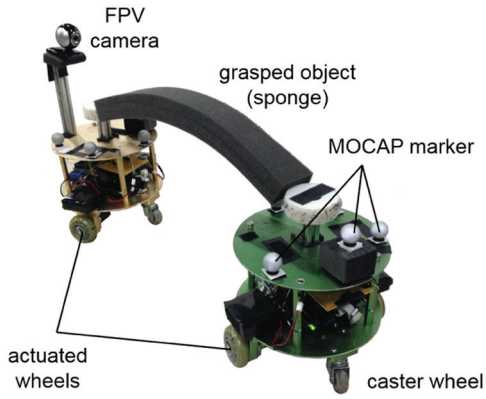

Fig. 1. Two nonholonomic WMRs cooperatively squeeze-grasping/transporting a deformable object (i.e. sponge) with no special grip-enforcing fixture and only via a passive part (i.e. rotating disk attached at the axle center), while being teleoperated via a FPV camera installed on the over-seeing WMR.

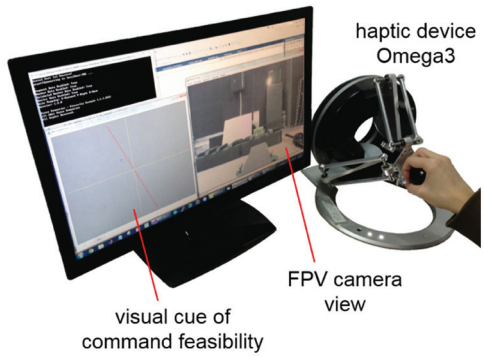

Fig. 2. Master interface with haptic device, FPV camera view and visual and haptic cues, notifying users how far and in which direction their command deviates from, and also nudging them onto the feasible motion set of the WMRs under the mixed constraint.

In this paper, we consider the teleoperation problem of two nonholonomic wheeled mobile robots (WMRs), which are cooperatively squeeze-grasping/transporting a deformable object without any dedicated (bilateral-contact) grip-enforcing fixture and only via a (unilateral-contact) passive part (i.e. rotating disk attached at the axle center), while being teleoperated by a single remote user through an FPV camera rigidly attached on the over-seeing (or following) WMR with a limited FOV (field of view). This FPV camera provides the user with information both on navigation and grasping aspects with the fore-running WMR always kept at the center of the FPV-FOV (i.e. FPV-centering). The human user tele-controls the forward and turning velocities of the FPV camera with some visuo-haptic feedback, which is to be designed to address the limited situation awareness of the FPV (see Figures 1 and 2).

To achieve the cooperative transport while maintaining the fixture-less grasping for the two nonholonomic WMRs is challenging. The key difficulty for this is that the two WMRs are, in this case, under a physical/nonholonomic constraint (i.e. no-slip/drift condition of the WMRs) and an artificial/holonomic constraint (i.e. fixture-less grasping and FPV-centering requirement) at the same time. This *mixed constraint* turns out to significantly curtail the possible motion space of the two WMRs so as to restrict the FPV camera is not able to track an arbitrary human command. This constrained FPV camera motion in turn complicates the teleoperation control design, since it is difficult for the human user to understand/predict how and why the FPV camera does not track their command only with the limited FPV, thereby, possibly inducing their confusion/frustration and consequently performance degradation. The teleoperation control design then needs to fully address this complication caused by the mixed constraint to still allow the users to tele-drive the WMRs intuitively and efficiently only via the limited FPV while also robustly enforcing the fixture-less grasping.

To circumvent this challenge, we utilize the nonoholonomic passive decomposition (Lee, 2010a,c) to first decompose the behavior (i.e. kinematics) of the two nonholonomic WMRs under the mixed constraint into the grasping/FPV-centering behavior and the FPV teleoperation behavior. Then, exploiting the "cognitive structure" of the problem (i.e. safety-critical/mathematically-simple grasping/FPV-centering, and error-tolerant/mathematically-vague tele-navigation), we propose a novel *FPV semi-autonomous teleoperation control architecture*, consisting of (low-level) autonomous grasp-enforcing and FPV-centering control and the (high-level) FPV camera teleoperation control, which in turn stands upon an autonomous human command tracking control. This semi-autonomous architecture then enables the human user to teleoperate the FPV camera, which will track the human command as much as permissible by the mixed constraint, while also guaranteeing the fixture-less grasping and FPV-centering (i.e. the fore-running WMR is always at the center of FPV-FOV) regardless of how aggressive/arbitrary the human command is. We also complement this semi-autonomous architecture with certain visuo-haptic feedback to notify the users of, and also nudge them to reduce, the command-behavior mismatch (i.e. if the FPV camera cannot fully track the human command) stemming from the mixed constraint. We also perform a human subject study to rigorously show that this visuo-haptic feedback indeed significantly improves the teleoperation performance and intuitiveness by substantially enhancing user perception/understanding of the command-behavior mismatch as compared to the case of using only a limited FPV.

In this paper, we choose differential-drive nonholonomic WMRs over omni-directional robots or mobile manipulators, since if they are successful they can result in a more affordable and rugged system, which is desirable in practice. Of course, this nonholonomic WMR has less dexterity than the systems mentioned above, which we believe would not be so limiting in many practical applications, since,

with the fixture-less grasping capability, the two WMRs can achieve such versatile operations as the handling of objects with various sizes/shapes only by squeezing and pushing. This is even more so if we deploy multiple nonholonomic WMRs, to which the theory presented in this paper can be easily extended.

We also mount the FPV camera on the over-seeing WMR to allow the user to supervise both the navigation and grasping aspects simultaneously (with fore-running WMR always at the center of the FPV-FOV by the FPV-centering control). Of course, we may attach multiple cameras on the WMRs, which is a topic for future research with its own issues and challenges to address (e.g. intuitive/efficient fusion of multiple views, real-time frame matching among (moving) multiple cameras, etc.). We also choose the two WMR system in Figure 1 since it is simple, yet, still rich enough, so that we can focus on the core issues of FPV semi-autonomous teleoperation with the obtained framework (e.g. decomposition-based control architecture) and findings (e.g. the possibility of replacing haptic feedback with visual feedback) which are applicable/extendable to more complex FPV teleoperation systems.

1.1. Contribution statement and related works

The contribution of this paper is three-fold:

(a) to propose a novel FPV semi-autonomous teleoperation control framework for robotic systems under the mixed constraint based on the behavior decomposition and the cognitive structure;
(b) to elucidate issues of using FPV for the teleoperation of robotic systems under a mixed constraint and to provide a systematic framework to design/validate visuo-haptic feedback to address this;
(c) to experimentally demonstrate the fixture-less cooperative grasping/transport of a deformable object only by using two nonholonomic WMRs under an arbitrary human (or object-level) command.

Numerous control schemes have been proposed for single mobile robot teleoperation (e.g. Buss et al., 2010; Ha et al., 2014; Karras et al., 2011; Lee and Xu, 2011; Lee et al., 2006; Nielsen et al., 2007). However, the peculiar issue of FPV associated with complex constrained motion as considered in this paper has not been explored, let alone the behavior decomposition, as the slave system itself is too simple to necessitate the issue.

Semi-autonomous teleoperation of multiple UAVs (unmanned aerial vehicles) was investigated by Franchi et al. (2012a,b) and Lee et al. (2013) with the efficacy of certain visuo-haptic feedback also studied by Son et al. (In Press), where a similar cognitive structure to ours (i.e. inter-agent formation and collective maneuver behaviors) was utilized. However, the behavior decomposition was not attained as their goal was only to loosely maintain the formation shape among the UAVs in contrast to the precise fixture-less grasping present in this paper. The issue of FPV teleoperation was also not considered in the work by Franchi et al. (2012a,b), Lee et al. (2013) and Son et al. (In Press).

A vision-based FPV semi-autonomous teleoperation scheme of a team of UAV-WMRs was proposed by Ha and Lee (2013), where a front-view FPV camera was used to navigate the team, whereas a down-view camera was used to maintain their formation with haptic feedback of the formation error. However, the scheme by Ha and Lee (2013) enforced the formation-keeping only loosely without navigation-formation behavior decomposition, thus, it was not capable of the rigid grasping-keeping required in this paper. Also related to this paper is asymmetric teleoperation (Malysz and Sirouspour, 2011a,b), where multiple users teleoperate a single slave robot. In this work, the null-space projection was used (i.e. prioritization) instead of complete behavior decomposition. Their slave system was also a single WMR or robotic manipulator, which is much simpler than our cooperative WMRs with no such mixed constraint issue as considered in our work. Further, the use of FPV and how to incorporate it into teleoperation in the form of visuo-haptic feedback was not addressed therein.

Our FPV semi-autonomous teleoperation framework proposed in this paper in fact extends its conference precursors (Lee, 2008; Lee and Spong, 2005), where similar behavior decomposition of multiple holonomic or nonholonomic WMRs were attained, yet their formation shape was restricted only to vectorial ones (i.e. object rotation prohibited), the issue of FPV not explored, and no experimental implementation and human subject study performed. To our knowledge, the problem of semi-autonomous teleoperation of slave systems under a mixed constraint with the issue of FPV fully incorporated into the teleoperation control design and human subject study to rigorously show its efficacy has not been addressed elsewhere and has been achieved in this paper for the first time.

Another major contribution of this paper is to experimentally demonstrate the fixture-less cooperative grasping/transport only by using two simple nonholonomic WMRs under arbitrary human (or object-level) command, which, to the best of our knowledge, has not been attained so far. The majority of results in this area either assumed physical (bilateral-contact) grip-enforcing fixtures (Alonso-Mora et al., 2015; Hirata et al., 2005; Khatib et al., 1996; Kim and Minor, 2010; Tang et al., 2006) or holonomic mobile manipulators (Alonso-Mora et al., 2015; Hirata et al., 2005; Khatib et al., 1996) to work around the issue of mixed constraint. A rare exception to this is the work by Sugar and Kumar (2002), where the fixture-less grasping was attained by multiple nonholonomic WMRs, but this required some extra specially-designed compliant grasping mechanisms to essentially separate the robots' locomotion from grasping, resulting in much more complex systems than our cooperative WMRs shown in Figure 1. We would emphasize that this contribution is from an "engineering"

perspective, as the theoretical enabler for this is the nonholonomic passive decomposition, which has already been proposed by (Lee (2010a,c).

The rest of this paper is organized as follows. Section 2 presents kinematic modeling of the cooperative nonholonomic WMRs and derives their nonholonomic passive decomposition under the mixed constraint according to the work by Lee (2010a,c). We then propose and detail a FPV semi-autonomous teleoperation control architecture in Section 3, which consists of autonomous fixture-less grasping and FPV-centering control and the FPV teleoperation control, which in turns relies on autonomous human command tracking control and is complemented with certain visuo-haptic feedback. Experimental verification and a human subject study for the proposed FPV semi-autonomous teleoperation framework and the efficacy of the visual and haptic feedback are given in Section 4. Section 5 contains some concluding remarks and comments on possible future research directions.

2. Kinematic modeling and decomposition of cooperative nonholonomic WMRs

In this section, we provide the kinematic modeling and analysis of the two nonholonomic WMRs, which are cooperatively grasping a common deformable object as shown in Figure 1. We find this kinematic modeling adequate for our purpose, with the operation speed of the WMRs not so fast and the ground friction large enough to prevent wheel slip/drift – see Section 4. As stated in Section 1, the two WMRs are under a mixed constraint, i.e. the physical/nonholonomic no-slip/drift constraint and the artificial/holonomic grasping/FPV-FOV requirement. To address this mixed constraint, we utilize the nonholonomic passive decomposition (Lee, 2010c,a,b). We also consider only the case of two WMRs and leave the case of multiple WMRs for future research, for which the theory and framework presented here can be extended in a straightforward manner.

2.1. Kinematic modeling of nonholonomic WMRs

The two WMRs we consider in this paper are differential-drive WMRs as shown in Figure 3 with the configuration $q_i := [x_i; y_i; \phi_i] \in \Re^3$, where (x_i, y_i) and ϕ_i are the axle center position and the yaw angle of the i-th WMR expressed in the inertial frame $\mathcal{I}$, $i = 1, 2$. It is then well-known that no slip/drift condition of the WMRs can be written by the Pfaffian constraint

$$A_i(q_i)\,\dot{q}_i = 0$$

with $A_i(q_i) := \begin{bmatrix} \sin\phi_i & -\cos\phi_i & 0 \end{bmatrix} \in \Re^{1\times 3}$.

Let us define the product system of the two WMRs by stacking their configurations s.t.

$$q := \begin{bmatrix} q_1; & q_2 \end{bmatrix} = \begin{bmatrix} x_1; & y_1; & \phi_1; & x_2; & y_2; & \phi_2 \end{bmatrix} \in \Re^6$$

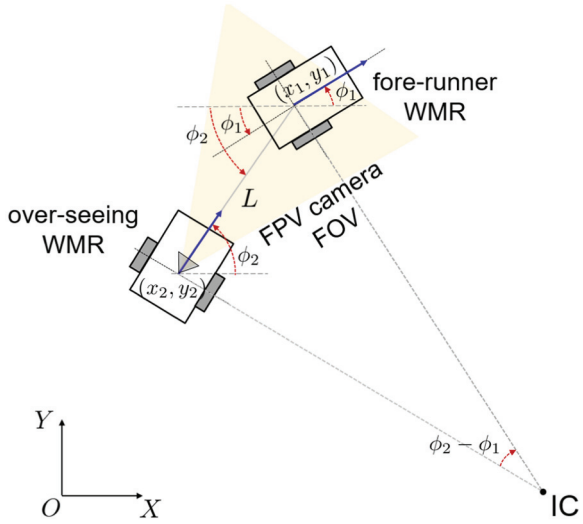

Fig. 3. Geometry of two nonholonomic WMRs when $h(q) = 0$, i.e. the distance between two axle centers of the WMRs is L (fixture-less grasping) with fore-running WMR being at the center of the FPV-FOV (FPV-centering). Also shown is the instantaneous center of rotation (ICR) of the two WMRs under $h(q) = 0$.

Then, the nonholonomic constraint for the product system can be written by

$$A(q)\,\dot{q} = 0 \tag{1}$$

with the constraint matrix $A(q) \in \Re^{2\times 6}$ given by

$$A(q) := \mathrm{diag}[A_1(q_1), A_2(q_2)] \in \Re^{2\times 6}$$

and the unconstrained distribution $\mathcal{D}^\top(q)$ (Lee, 2010c), which characterizes the sub-space of velocity respecting the nonholonomic constraint given by equation (1), can be identified by the following matrix

$$\mathcal{D}_\top(q) := \begin{bmatrix} \mathrm{c}\,\phi_1 & \mathrm{s}\,\phi_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{c}\,\phi_2 & \mathrm{s}\,\phi_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T \in \Re^{6\times 4} \tag{2}$$

where $\mathrm{c}\,\phi_i = \cos\phi_i$ and $\mathrm{s}\,\phi_i = \sin\phi_i$. Here, note that $\mathcal{D}_\top$ identifies the null-space of $A(q) \in \Re^{2\times 6}$ in equation (1).

Evolution of the two WMRs under the nonholonomic constraint given by equation (1) can then be written by the following drift-less nonlinear control equation

$$\dot{q} = \mathcal{D}_\top(q)\,u \tag{3}$$

where $u := [v_1; w_1; v_2; w_2] \in \Re^4$ is the control input, with $v_i, w_i \in \Re$ being the forward and angular velocity commands of the i-th WMR. We can show that the constraint given by equation (1) is completely nonholonomic, with the closure of Lie brackets (Lee, 2010c; Murray et al., 1993) of all the columns of $\mathcal{D}_\top(q)$ spans $\Re^6$, suggesting that, by combining v_i, w_i, the two WMRs can assume any configuration $q \in \Re^6$.

2.2. Design of the grasping shape map $h(q)$

We aim to achieve fixture-less grasping and transport, that is, only through the (unilateral contact) passive part installed at their axle center as shown in Figure 1, the two WMRs should be able to squeeze the deformable object and maintain this grasping during the transport without any (bilateral-contact) grip-enforcing fixture. This then requires the two WMRs to maintain a certain distance $L > 0$ between their axle centers throughout the operation. On the other hand, a remote human user should be able to tele-drive the two cooperative WMRs only by looking through the FPV camera. This FPV camera is installed on the over-seeing WMR to provide the user with information on both the navigation and grasping aspects, so that they can respond to, for example, the fall of the grasped object in unexpected accidents while tele-navigating the two WMRs. The FPV camera, however, possesses only a limited FOV, and it is necessary to maintain the fore-running WMR always within the limited FOV of the FPV camera. One way to achieve this is to control the WMRs in such a way that the fore-running WMR is always at the center of the FPV-FOV (see Figure 3).

These two requirements, i.e. fixture-less grasping and FPV-centering, can be written as the following artificial/holonomic constraint

$$h(q) := \begin{pmatrix} x_1 - x_2 - L\,\mathrm{c}\,\phi_2 \\ y_1 - y_2 - L\,\mathrm{s}\,\phi_2 \end{pmatrix} = 0 \tag{4}$$

where $L > 0$ is the desired squeezing distance. Note that, if $h(q) = 0$, the distance between the two axle centers of the WMRs is maintained to be L (i.e. fixture-less grasping) while the over-seeing WMR is always pointing to the fore-running WMR (i.e. FPV-centering, see Figure 3). This $h(q) = 0$ is an *artificial* constraint (i.e. control objective), which needs to be attained/maintained by the autonomous control of Section 3.1. This $h(q)$ is called a *formation map* or *grasping shape map* (Lee, 2010c). Here, we assume $|\phi_1(t) - \phi_2(t)| < \pi \ \forall t \geq 0$, which is ensured by the autonomous control as stated in Section 3.1 and can also be seen from Section 4.

Following (Lee and Li, 2013), we can then define the *tangential distribution* $\Delta^{\top}(q)$ of $h(q)$ to be the null-space of the one-form $\frac{\partial h}{\partial q} \in \Re^{2\times 6}$ as identified by the following matrix

$$\Delta_{\top}(q) = \begin{bmatrix} L\,\mathrm{s}\,\phi_2 & -L\,\mathrm{c}\,\phi_2 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}} \in \Re^{6\times 4} \tag{5}$$

and the *normal distribution* $\Delta^{\perp}(q)$, which is the orthogonal complement of $\Delta^{\top}$ with respect to the Euclidean metric as identified by

$$\Delta_{\perp}(q) = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & L\,\mathrm{s}\,\phi_2 \\ 0 & 1 & 0 & 0 & -1 & -L\,\mathrm{c}\,\phi_2 \end{bmatrix}^{\mathrm{T}} \in \Re^{6\times 2} \tag{6}$$

with $\Delta_{\perp}(q) = \left(\frac{\partial h}{\partial q}\right)^{\mathrm{T}}$. As explained by Lee and Li (2013), the velocity component in $\Delta^{\top}$ characterizes the motion of the two WMRs tangential to the (current) level set

$$\mathcal{H}_{h(q)} := \{q' \in \Re^6 \mid h(q') = h(q)\} \tag{7}$$

(i.e. motion keeping the grasping shape $h(q)$ intact); whereas that in $\Delta^{\perp}$ the motion normal to the level set $\mathcal{H}_{h(q)}$ with respect to the Euclidean metric (i.e. in the direction of changing $h(q)$).

2.3. Nonholonomic passive decomposition

As shown above, the two cooperative WMRs are under a mixed constraint, i.e. a physical/nonholonomic constraint (i.e. unconstrained distribution $\mathcal{D}^{\top}$) and an artificial/holonomic requirement (i.e. tangential and normal distributions $\Delta^{\top}, \Delta^{\perp}$). To facilitate the analysis and control synthesis for two WMRs under this mixed constraint, we utilize the kinematic version of the nonholonomic passive decomposition (Lee, 2010a,c). Readers who do not want to go through the detailed derivation of this decomposition may go directly go to equations (12) and (13). We also refer readers to the work by Lee (2010a,b,c) for more details and explanations of the nonholonomic passive decomposition.

The nonholonomic passive decomposition of general robotic systems is given by Lee (2010a,c)

$$\mathcal{D}^{\top} = (\mathcal{D}^{\top} \cap \Delta^{\top}) \oplus (\mathcal{D}^{\top} \cap \Delta^{\perp}) \oplus \mathcal{D}^{c} \tag{8}$$

where $(\mathcal{D}^{\top} \cap \Delta^{\top})$ and $(\mathcal{D}^{\top} \cap \Delta^{\perp})$ are called (unconstrained) *locked and shape distributions*, respectively representing the tangential (i.e. $h(q)$ locked) and normal (i.e. $h(q)$ changing) components of the level set $\mathcal{H}_{h(q)}$ equation (7) among the permissible robot motion in $\mathcal{D}^{\top}$ (i.e. satisfying nonholonomic constraint). The distribution $\mathcal{D}^{c}$ is called the *quotient distribution*, which is still permissible, yet contains both the tangential and normal components of the level set $\mathcal{H}_{h(q)}$ and cannot be split either into a tangential or normal direction only. As will be shown below, for the cooperative WMRs, $\mathcal{D}^{c} \neq \emptyset$ implying weak decomposability (Lee, 2010c).

We can then compute the matrices identifying each distribution in equation (8) for the two WMRs given by equation (3) under the grasping map $h(q)$ in equation (4) as follows. First, we can immediately notice from equations (2) and (5) that

$$\mathrm{span}\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \subset (\mathcal{D}^{\top} \cap \Delta^{\top})$$

We can also find the remaining components of $(\mathcal{D}^{\top} \cap \Delta^{\top})$ by solving the next equation for $\nu_1 = [a_1; b_1; c_1] \in \Re^3$

$$\left[A \cdot \Delta_{\top}^{1:3}\right] \nu_1 = \begin{bmatrix} L\,\mathrm{c}(\phi_1 - \phi_2) & \mathrm{s}\,\phi_1 & -\mathrm{c}\,\phi_1 \\ 0 & \mathrm{s}\,\phi_2 & -\mathrm{c}\,\phi_2 \end{bmatrix} \nu_1 = 0$$

where $\Delta_{\top}^{1:3} \in \Re^{6\times 3}$ is the matrix formed from the first three columns of $\Delta_{\top} \in \Re^{6\times 4}$. We then see that

rank $\left[A(q)\cdot\Delta_{\top}^{1:3}(q)\right] = 2\ \forall q \in \Re^6$, since this rank can be less than two only if $\mathrm{c}(\phi_1-\phi_2)=0$. However, for this case, the determinant of the right-most two-by-two sub-matrix of $\left[A\cdot\Delta_{\top}^{1:3}\right]$ is $\mathrm{s}(\phi_1-\phi_2)\neq 0$. This then means that rank $\left[\mathcal{D}^{\top}\cap\Delta^{\top}\right]=2$ and the distribution $\left[\mathcal{D}^{\top}\cap\Delta^{\top}\right]$ can be identified by the following matrix

$$(\mathcal{D}_{\top}\cap\Delta_{\top}) = \begin{bmatrix} \mathrm{c}\,\phi_1 & 0 \\ \mathrm{s}\,\phi_1 & 0 \\ 0 & 1 \\ \mathrm{c}(\phi_1-\phi_2)\,\mathrm{c}\,\phi_2 & 0 \\ \mathrm{c}(\phi_1-\phi_2)\,\mathrm{s}\,\phi_2 & 0 \\ \frac{1}{L}\,\mathrm{s}(\phi_1-\phi_2) & 0 \end{bmatrix} \in \Re^{6\times 2} \tag{9}$$

We can similarly compute $(\mathcal{D}^{\top}\cap\Delta^{\perp})$, that is, by solving the following equation for $\nu_2=[a_2;b_2]\in\Re^2$

$$[A\cdot\Delta_{\perp}]\,\nu_2 = \begin{bmatrix} \mathrm{s}\,\phi_1 & -\mathrm{c}\,\phi_1 \\ -\mathrm{s}\,\phi_2 & \mathrm{c}\,\phi_2 \end{bmatrix}\nu_2 = 0$$

which does not have any solution unless $\mathrm{s}(\phi_1-\phi_2)=0$, i.e. $\phi_1=\phi_2$, with $|\phi_1-\phi_2|<\pi$ assumed as stated after equation (4). This then means that the distribution $(\mathcal{D}^{\top}\cap\Delta^{\perp})$ can be identified by

$$(\mathcal{D}_{\top}\cap\Delta_{\perp}) = \tag{10}$$
$$\begin{cases} \emptyset & \text{if } \phi_1\neq\phi_2 \\ \begin{bmatrix} \mathrm{c}\,\phi_1 & \mathrm{s}\,\phi_1 & 0 & -\mathrm{c}\,\phi_2 & -\mathrm{s}\,\phi_2 & 0 \end{bmatrix}^{\mathrm{T}} & \text{if } \phi_1=\phi_2 \end{cases}$$

On the other hand, we can obtain the matrix for the quotient distribution $\mathcal{D}^c$ as follows. First, let us compute $\mathcal{D}^{\top}\backslash(\mathcal{D}^{\top}\cap\Delta^{\top})$ by solving $\nu_3=[a_3;b_3;c_3]\in\Re^3$ for the equation below

$$[(\mathcal{D}_{\top}\cap\Delta_{\top})_1^{\mathrm{T}}\cdot\mathcal{D}_{\top}^{1,3:4}]\nu_3 = \begin{bmatrix} 1 & \mathrm{c}(\phi_1-\phi_2) & \frac{1}{L}\,\mathrm{s}(\phi_1-\phi_2) \end{bmatrix}\nu_3 = 0$$

where $(\mathcal{D}_{\top}\cap\Delta_{\top})_1$ is the first column vector of $(\mathcal{D}_{\top}\cap\Delta_{\top})$ in equation (9) and $\mathcal{D}_{\top}^{1,3:4}\in\Re^{6\times 3}$ is the matrix formed from the first, third and fourth columns of $\mathcal{D}_{\top}$ in equation (2), as it is obvious that span $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ is contained both in $\mathcal{D}^{\top}$ and $\mathcal{D}^{\top}\cap\Delta^{\top}$. The solution of this equation does in fact identify the matrix $\mathcal{D}_c$ for the quotient distribution $\mathcal{D}^c$, i.e.

$$\mathcal{D}_{\top}\backslash(\mathcal{D}_{\top}\cap\Delta_{\top}) = \mathcal{D}_c \tag{11}$$
$$= \begin{bmatrix} \mathrm{c}\,\phi_1 & 0 \\ \mathrm{s}\,\phi_1 & 0 \\ 0 & 0 \\ -\mathrm{c}(\phi_1-\phi_2)\,\mathrm{c}\,\phi_2 & \mathrm{s}(\phi_1-\phi_2)\,\mathrm{c}\,\phi_2 \\ -\mathrm{c}(\phi_1-\phi_2)\,\mathrm{s}\,\phi_2 & \mathrm{s}(\phi_1-\phi_2)\,\mathrm{s}\,\phi_2 \\ -L\,\mathrm{s}(\phi_1-\phi_2) & -L\,\mathrm{c}(\phi_1-\phi_2) \end{bmatrix} \in \Re^{6\times 2}$$

since $\Delta_{\top}^{\mathrm{T}}\cdot[\mathcal{D}_{\top}\backslash(\mathcal{D}_{\top}\cap\Delta_{\top})]\neq 0$ and $\Delta_{\perp}^{\mathrm{T}}\cdot[\mathcal{D}_{\top}\backslash(\mathcal{D}_{\top}\cap\Delta_{\top})]\neq 0$, that is $\mathcal{D}^{\top}\backslash(\mathcal{D}^{\top}\cap\Delta^{\top})$, is not contained in either $\Delta^{\top}$ or $\Delta^{\perp}$. Note also that the first column of $\mathcal{D}_c$ in equation (11) subsumes $(\mathcal{D}_{\top}\cap\Delta_{\perp})$ when $\phi_1=\phi_2$.

Using the matrices identifying each distribution of equation (8) as derived above, we can then rewrite the kinematics of the two WMRs given by equation (3) s.t.

$$\dot{q} = \underbrace{\begin{bmatrix} (\mathcal{D}_{\top}\cap\Delta_{\top}) & \mathcal{D}_c \end{bmatrix}}_{=:S(q)\in\Re^{6\times 4}} \begin{pmatrix} u_L \\ u_C \end{pmatrix} \tag{12}$$

where $u_L=[u_L^1;u_L^2]\in\Re^2$ and $u_C=[u_C^1;u_C^2]\in\Re^2$ are respectively the velocity command of the locked system, which describes the motion of the two WMRs along the direction of keeping $h(q)$ locked (e.g. fixture-less grasping and FPV-centering maintained), and that of the quotient system, which can incur the change of the grasping shape map $h(q)$, as $\mathcal{D}^c$ contains some components in $\Delta^{\perp}$. The matrix $S(q)$ is called the decomposition matrix. This kinematic equation (12) is the basis for our control design in the next section. Once we design the locked and quotient controls u_L and u_C in Section 3, we can decode them to the original control u in equation (3) s.t.

$$u = \left[\mathcal{D}_{\top}^T(q)\,\mathcal{D}_{\top}(q)\right]^{-1}\mathcal{D}_{\top}^T(q)\,S(q)\begin{pmatrix} u_L \\ u_C \end{pmatrix} \tag{13}$$

which can be obtained directly from equation (3). Refer to the work by Lee (2010c,a,b) for more details on the nonholonomic passive decomposition and some control techniques associated with this.

3. Semi-autonomous teleoperation control architecture design

In this section, we design a FPV semi-autonomous teleoperation control for the two WMRs given by equation (1) under the grasping shape map requirement (equation (4)) based on their nonholonomic passive decomposition (equation (12)). Our proposed control architecture consists of two control loops.

1. The (low-level) autonomous grasping/FPV-centering control, which enforces $h(q)=0$ to attain the fixture-less grasping/transport while keeping the fore-running WMR at the center of the FPV-FOV regardless of the human command.
2. The (high-level) teleoperation control, which drives the FPV camera to track the human command (i.e. forward velocity and turning rate) as much as is permissible by the mixed constraint with some visuo-haptic feedback to notify the user of, and nudge them to reduce, the command-behavior mismatch. This is difficult for users to understand/predict only with the limited FPV, thus may cause their confusion/frustration and substantial performance degradation as can be seen from Section 4.

3.1. Autonomous grasping and FPV-centering control

We first design the autonomous fixture-less grasping and FPV-centering control, whose objective can be written by a simple mathematical expression (i.e. $h(q)=0$ in equation (4)), which can be fully-automated. For this, we design u_C in equation (12) s.t.

$$u_C = -\mathcal{D}_c^{\mathrm{T}}(q)\cdot\left[\frac{\partial h}{\partial q}\right]^{\mathrm{T}}\left[\frac{\partial \varphi_h}{\partial h}\right]^{\mathrm{T}} \in \Re^2 \qquad (14)$$

where $\varphi_h(h)\geq 0$ is a suitably defined positive-definite potential function to enforce $h(q)=0$ (e.g. with $\varphi_h(h)=0$ if and only if $h=0$ and is positively quadratic around the neighborhood of $h=0$). Here, note that $\frac{\partial h}{\partial q}\in\Re^{2\times 6}$ and $\frac{\partial \varphi_h}{\partial h}\in\Re^{2\times 2}$. We then have

$$\frac{d\varphi_h}{dt}=\frac{\partial\varphi_h}{\partial h}\frac{\partial h}{\partial q}\dot{q}=\frac{\partial\varphi_h}{\partial h}\frac{\partial h}{\partial q}\mathcal{D}_c u_C=-||u_C||^2\leq 0$$

where we use the fact that $\frac{\partial h}{\partial q}\cdot[\mathcal{D}_\top\cap\Delta_\top]\cdot u_L=0$ regardless of u_L in equation (13) from the construction of $\Delta^\top$. Note that this u_L will embed the teleoperation control from Section 3.2.2.

This then means that $\varphi_h(t)\leq\varphi_h(0)$ and $\forall t\geq 0$, that is, if the grasping and FPV-centering error starts small (i.e. $|h(q(0))|\approx 0$), it will stay small (i.e. $|h(q(t))|\approx 0$ $\forall t\geq 0$). Further, this error $\varphi_h(h)$ will be strictly decreasing as long as $u_C\neq 0$. This also implies that, if we start with a small enough $h(q(0))$ with $|\phi_1(0)-\phi_2(0)|<\pi$, $|\phi_1(t)-\phi_2(t)|<\pi$ $\forall t\geq 0$ as stated after equation (4), since if not, $\varphi_h(t)$ should increase contradictory to the fact that $\varphi_h(t)\leq\varphi_h(0)$, $\forall t\geq 0$. Here, note that:

(a) this fixture-less grasping and the FPV-centering are enforced regardless of the teleoperation control (i.e. u_L in equation (13));
(b) conversely, this autonomous grasping/FPV-centering control u_C does not at all affect the teleoperation aspect (i.e. $(\mathcal{D}^\top\cap\Delta^\top)$ in equation (12)) with $(\mathcal{D}_\top\cap\Delta_\top)^{\mathrm{T}}\cdot\mathcal{D}_c\cdot u_C=0$, $\forall u_C\in\Re^2$.

This grasping-teleoperation decoupling is due to the split of their behaviors through the nonholonomic passive decomposition given by equation (12).

3.2. Teleoperation control with visuo-haptic feedback

Here, we design the teleoperation control to allow a user to remotely drive the two WMRs via the FPV camera under a mixed constraint. In contrast to the autonomous grasping/FPV-centering control from Section 3.1, which is safety-critical, yet mathematically simple (i.e. $h(q)=0$ in equation (4)) and so can be fully-automated, this tele-driving is more error-tolerant, yet mathematically-vague/evasive to be captured by such a simple mathematical expression, particularly when the task environment is unknown, uncertain or dynamic. For this, human intelligent intervention and cognitive decision are necessary.

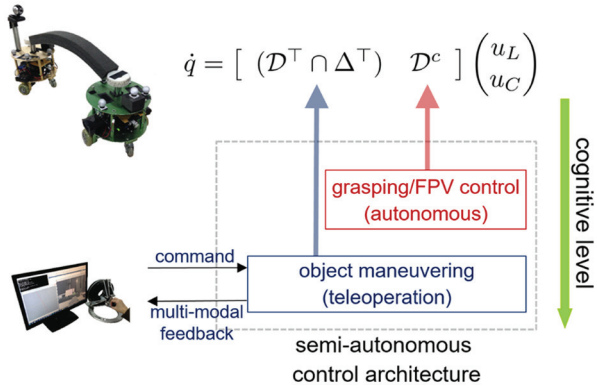

Fig. 4. FPV semi-autonomous teleoperation architecture for the cooperative WMRs, consisting of: 1) (low-level) autonomous fixture-less grasping and FPV-centering control with safety-critical/mathematically-simple objective (i.e. $h(q)=0$); and 2) (high-level) FPV teleoperation with an error-tolerant/mathematically-vague objective with visuo-haptic feedback of command-behavior mismatch.

This cognitive structure then suggests the *semi-autonomous teleoperation architecture* shown in Figure 4, consisting of the (low-level) local autonomous control layer to address simple yet "hard/safety-critical" requirements (i.e. fixture-less grasping and FPV-centering with $h(q)=0$), and the (high-level) teleoperation control layer to allow the human user to tele-control "soft/decisional" aspects of the operation (i.e. tele-navigation of the two WMRs). We also complement this semi-autonomous architecture with visuo-haptic feedback to address the command-behavior mismatch, which inevitably arises due to the mixed constraint, but is difficult for the user to understand only with the limited FPV. Thus, this can significantly deteriorate intuitiveness and the performance of teleoperation if not treated properly (see Section 4).

3.2.1. Restricted FPV camera motion under mixed constraint. To better see this command-behavior mismatch, we need to understand how the FPV camera motion is restricted. For this, consider the kinematic equation of the two WMRs (equation (12)), where $u_C\in\Re^2$ has already been used up to enforce the artificial/holonomic constraint $h(q)=0$ as designed in equation (14). Recall also that the FPV camera is attached on the over-seeing WMR. Then, from equation (12) with equation (9), we can see that this over-seeing WMR now has only direct control available (i.e. $u_L^1\in\Re$), implying that the FPV in general cannot track an arbitrary human command, which has two-DOF (i.e. the forward velocity and turning rate of the FPV), particularly given that precise prediction of human intention (and its feedforward action) is generally impossible.

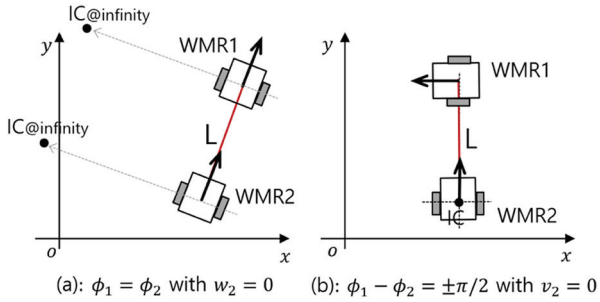

Fig. 5. Instantaneous center of rotation (ICR) and singular cases when $\phi_1 = \phi_2$ (with the ICR at infinity) and $\phi_1 - \phi_2 = \pm\pi/2$ under the grasping shape requirement $h(q) = 0$ in equation (4).

This restricted FPV camera motion can also be understood by using the notion of instantaneous center of rotation (ICR) as depicted in Figure 3. That is, if we enforce $h(q) = 0$ (with $u_C = 0$ as well from equation (14)), the motion of the two WMRs becomes constrained as if they hold a rigid object of the length L with the over-seeing WMR always fixed toward the center of the fore-running WMR, while this fore-running WMR can freely rotate. This then means that the collective of the two WMRs under $h(q) = 0$ has only two-DOF, i.e.:

(a) the collective rotation as if it is one single rigid-body about the ICR with the angular rate of $\dot{\phi}_2 = \frac{1}{L}\,\mathrm{s}(\phi_1 - \phi_2)\,u_L^1$;
(b) the fore-running self-rotation of the WMRs with $\dot{\phi}_1 = u_L^2$, with the first and second columns of equation (9) respectively corresponding to these.

See also Figure 5 for the two singular cases:

(a) no rotation of the over-seeing WMR is possible (i.e. $\dot{\phi}_2 = 0$) when $\phi_1 = \phi_2$;
(b) no forward velocity of the over-seeing WMR is possible (i.e. $v_2 := u_L^1\,\mathrm{c}(\phi_1 - \phi_2) = 0$) when $\phi_1 - \phi_2 = \pm\pi/2$.

Note also from equation (9) that these two singular cases (i.e. $v_2 = 0$ or $\dot{\phi}_2 = 0$) cannot happen at the same time.

Due to this restriction and also with the impossibility of accurate human intention prediction and limitation of feedback control gain in practice, the FPV camera in general cannot precisely track arbitrary human command. To address this, we will design the human command tracking control (Section 3.2.2) so that the FPV camera tracks the human command as much as permissible by the mixed constraint, while notifying the user of, and also nudging them to reduce, the command-behavior mismatch via visual and haptic feedback (Section 3.2.3). This turns out to significantly improve intuitiveness/performance of teleoperation as compared to the case of only with a limited FPV (Section 4).

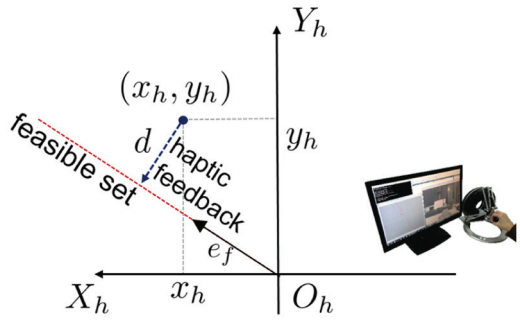

Fig. 6. Master device space for FPV teleoperation, where the position (x_h, y_h) of the haptic device is used to command the turning and forward velocities of the FPV camera. Also shown is the feasible set mapped from Figure 7, command-behavior mismatch $d \in \Re^2$, and the visual/haptic feedback.

3.2.2. Human command tracking control. For the teleoperation, we assume the master device has two-DOF to command the forward and turning velocities of the FPV camera. One example is shown in Figure 6, where the position $(x_h, y_h) \in \Re$ of the haptic device is used with the following command mapping

$$\begin{pmatrix} v_2^d \\ w_2^d \end{pmatrix} := \begin{pmatrix} \eta_1 \cdot y_h \\ \eta_2 \cdot x_h \end{pmatrix} \tag{15}$$

where v_2^d and w_2^d are the desired forward/turning velocities of the over-seeing WMR (and FPV camera) with $\eta_1, \eta_2 > 0$ being some scaling factors to map the master configuration to the velocity of the over-seeing WMR, thereby overcoming the master-slave kinematic dissimilarity via the position-velocity mapping (Lee, 2008; Lee and Spong, 2005; Lee et al., 2013, 2006). Here, note from Figure 3 that positive x_h (i.e. move to left) corresponds to positive $\dot{\phi}_2$ (i.e. view rotating to left).

For the human command tracking, we then want the following: from equation (12) with equation (9)

$$v_2' := \cos(\phi_1 - \phi_2)\cdot u_L^1 \to v_2^d \tag{16}$$

$$w_2' := \frac{1}{L}\sin(\phi_1 - \phi_2)\cdot u_L^1 \to w_2^d \tag{17}$$

where (v_2', w_2') reduces to the forward/turning velocities of the over-seeing WMR (v_2, w_2), when $h(q) = 0$ (with $u_C = 0$ as well from equation (14)), which is to be enforced by the autonomous grasping/FPV-centering control designed in Section 3.1. Now, as stated in Section 3.2.1, for the two control objectives given by equations (16) and (17), we only have one direct control u_L^1. Yet, from equation (9), we can also see that the angle of the fore-running WMR ϕ_1 can be directly controlled by u_L^2. This then suggests the use of ϕ_1 and u_L^1 as the control inputs to attain equations (16) and (17), similar to the idea of backstepping (Ha et al., 2014; Sepulchre et al., 1997) and dynamic extension (Kolmanovsky and McClamroch, 1995).

More specifically, let us define the following Lyapunov function

$$V := \frac{1}{2}e_v^2 + \frac{1}{2}e_w^2 \tag{18}$$

where $e_v := v_2' - v_2^d$ and $e_w := w_2' - w_2^d$. Then, using equation (12) and equation (9), we have

$$\begin{aligned}\frac{dV}{dt} =& e_v[\mathrm{c}(\phi_1-\phi_2)\,\dot{u}_L^1 - \mathrm{s}(\phi_1-\phi_2)\,(\dot{\phi}_1-\dot{\phi}_2)\,u_L^1 - \dot{v}_2^d] \\ &+ e_w[\tfrac{1}{L}\,\mathrm{s}(\phi_1-\phi_2)\,\dot{u}_L^1 + \tfrac{1}{L}\,\mathrm{c}(\phi_1-\phi_2)\,(\dot{\phi}_1-\dot{\phi}_2)\,u_L^1 - \dot{w}_2^d] \\ =& \begin{pmatrix} e_v \\ e_w \end{pmatrix}^{\mathrm{T}} \left[P\begin{pmatrix} \dot{u}_L^1 \\ u_L^2 \end{pmatrix} + \begin{pmatrix} \mathrm{s}(\phi_1-\phi_2)\,\dot{\phi}_2 u_L^1 \\ -\frac{1}{L}\,\mathrm{c}(\phi_1-\phi_2)\,\dot{\phi}_2 u_L^1 \end{pmatrix} - \begin{pmatrix} \dot{v}_2^d \\ \dot{w}_2^d \end{pmatrix} \right]\end{aligned}$$

where

$$P := \begin{bmatrix} \mathrm{c}(\phi_1-\phi_2) & -\mathrm{s}(\phi_1-\phi_2)\,u_L^1 \\ \frac{1}{L}\,\mathrm{s}(\phi_1-\phi_2) & \frac{1}{L}\,\mathrm{c}(\phi_1-\phi_2)\,u_L^1 \end{bmatrix} \tag{19}$$

and we also use the fact that $\dot{\phi}_1 = u_L^2$ from the structure of equation (12) with equations (9) and (11), regardless of the autonomous grasping/FPV-centering control u_C in equation (14).

We then design the control to achieve equations (16) and (17) to be

$$\begin{pmatrix} \dot{u}_L^1 \\ u_L^2 \end{pmatrix} = P^{-1}\left[\begin{pmatrix} \dot{v}_2^d \\ \dot{w}_2^d \end{pmatrix} - \begin{pmatrix} \mathrm{s}(\phi_1-\phi_2)\,\dot{\phi}_2 u_L^1 \\ -\frac{1}{L}\,\mathrm{c}(\phi_1-\phi_2)\,\dot{\phi}_2 u_L^1 \end{pmatrix} - \Lambda\begin{pmatrix} e_v \\ e_w \end{pmatrix}\right] \tag{20}$$

where $\Lambda \in \Re^{2\times 2}$ is a positive-definite and symmetric gain matrix. Then, with this control equation (20), we have

$$\frac{dV}{dt} = -\begin{pmatrix} e_v \\ e_w \end{pmatrix}^{\mathrm{T}} \Lambda \begin{pmatrix} e_v \\ e_w \end{pmatrix}$$

implying that $(e_v, e_w) \to 0$ exponentially, which also suggests robustness against model uncertainty as well. We can then analytically solve equation (20) to obtain u_L^2; whereas we can integrate equation (20) to compute u_L^1, given $\phi_1 - \phi_2$, $\dot{\phi}_2$ and u_L^1. Once this $u_L = (u_L^1, u_L^2)$ is obtained, the original control $u \in \Re^4$ of equation (3) can be retrieved via the decoding equation (13) with the autonomous grasping/FPV-centering control $u_C \in \Re^2$, as designed in equation (14).

The control equation (20) becomes singular when $u_L^1 = 0$, since

$$\det P = \frac{1}{L}[\mathrm{c}^2(\phi_1-\phi_2) + \mathrm{s}^2(\phi_1-\phi_2)]u_L^1 = \frac{1}{L}u_L^1$$

The situation of $u_L^1 = 0$ cannot be avoided for the teleoperation though, as the human command is in general arbitrary and unpredictable. To address this, we utilize a bootstrapping-like strategy as follows. First, note that $u_L^1 = 0$ may occur when the human user indeed wants to stop the two WMRs with $v_2^d = w_2^d = 0$ for equations (16) and (17). For this case, it is not desired to disturb $u_L^1 = 0$. On the other hand, $u_L^1 = 0$ may occur incidentally in the course of integrating equation (20), although $v_2^d, w_2^d \neq 0$. In this case, it would be necessary to perturb $u_L^1 = 0$ with some non-zero bootstrapping value $\bar{u}_L^1$. This perturbation should be consistent with the human teleoperation command (or intention) v_2^d, w_2^d.

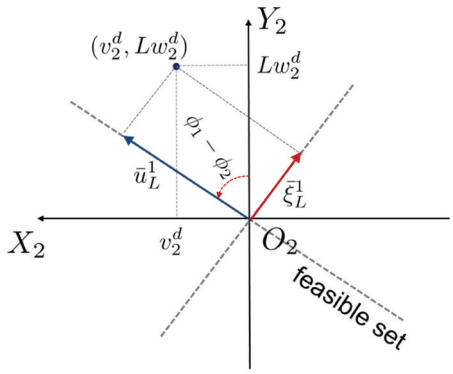

Fig. 7. The (scaled) human command space with coordinate (v_2^d, Lw_2^d), feasible set according to equations (16) and (17) and orthogonal coordinate transformation $(\bar{u}_L^1, \bar{\xi}_L^1)$ along the feasible and infeasible directions.

To determine this bootstrapping $\bar{u}_L^1$ while incorporating the user intention, consider the space of the (scaled) human command with the coordinate (v_2^d, Lw_2^d) as shown in Figure 7. We also say that the human command (v_2^d, w_2^d) is feasible, if, given ϕ_1, ϕ_2, $\exists\, u_L^1 \in \Re$ s.t.

$$\begin{aligned}\cos(\phi_1-\phi_2)\cdot u_L^1 &= v_2^d \\ \tfrac{1}{L}\sin(\phi_1-\phi_2)\cdot u_L^1 &= w_2^d\end{aligned}$$

i.e. this human command (v_2^d, w_2^d) can be instantaneously generated by the over-seeing WMR (with $h(q) = 0$) even if it assumes only one control u_L^1. Here, note that the human command (v_2^d, w_2^d) is in general arbitrary, thus, not necessarily always feasible. Even if $\det P \neq 0$, the tracking performance of equations (16) and (17) can never be perfect either in practice, since the human command is unpredictable (i.e. feedforward is impossible) and the implementation is always imperfect in practice with e.g. limited control gain Λ, sensing/actuation delay, sampling and quantization, etc.

We can then define the following coordinate transformation of (v_2^d, Lw_2^d) s.t.

$$\begin{pmatrix} \bar{u}_L^1 \\ \bar{\xi}_L^1 \end{pmatrix} = \begin{bmatrix} \mathrm{c}(\phi_1-\phi_2) & \mathrm{s}(\phi_1-\phi_2) \\ \mathrm{s}(\phi_1-\phi_2) & -\mathrm{c}(\phi_1-\phi_2) \end{bmatrix}\begin{pmatrix} v_2^d \\ Lw_2^d \end{pmatrix} \tag{21}$$

where $\bar{u}_L^1 \in \Re$ denotes the feasible component of the command (v_2^d, Lw_2^d); whereas $\bar{\xi}_L^1$ is the component not permissible (or infeasible) by the mixed constraint (i.e., command-behavior mismatch). We call the line of $\bar{u}_L^1$ *feasible set* in the (v_2^d, Lw_2^d)-space (see Figure 7). The feasible component $\bar{u}_L^1$ can then be computed by

$$\bar{u}_L^1 = v_2^d\,\mathrm{c}(\phi_1-\phi_2) + Lw_2^d\,\mathrm{s}(\phi_1-\phi_2) \tag{22}$$

which is in fact the optimal projection of (v_2^d, Lw_2^d) into the feasible set with respect to the Euclidean metric.

We then replace u_L^1 in equation (20) with this $\bar{u}_L^1$ given by equation (22) whenever $u_L^1(t) = 0$ (i.e. $u_L^1(t) \leftarrow \bar{u}_L^1(\phi_1 - \phi_2, v_2^d, w_2^d)$) and solve/integrate equation (20) to compute u_L^1, u_L^2. It may also happen that $u_L^1(t) = 0$, yet $\bar{u}_L^1(t) = 0$ as well (i.e. human command $(v_2^d, w_2^d) = 0$ or completely infeasible purely along the line of ξ_L^1). For this, we set $(u_L^1(t), u_L^2(t)) \leftarrow 0$ to stop the two WMRs, since, in this case:

(a) the user may explicitly command to stop the operation with $(v_2^d, w_2^d) = 0$ or;
(b) implicitly suggest so with their command completely infeasible even if the visual or haptic feedback keeps notifying them that their command is completely infeasible, thus, the FPV camera may not track their command satisfactorily.

Of course, this inference may be incorrect, yet in this case, the user will anyway produce non-zero $\bar{u}_L^1$ again at the next step, with which the integration/computation of equation (20) and also the operation of the two WMRs will resume.

3.2.3. Visual and haptic feedback design. Although, in theory, precise sensing of $\dot{v}_2^d = \eta_1 \cdot \dot{y}_h$ and $\dot{w}_2^d = \eta_2 \cdot \dot{x}_h$ from the master device (equation (15)) may be assumed, in practice, there will always be error/lag in estimating this human command $(\dot{v}_2^d, \dot{w}_2^d)$ due to, e.g. encoder quantization, sampling effect, etc. Human intention cannot be accurately predicted either, meaning that feedforward action to eliminate such sensing inaccuracy is impossible. The feedback gain Λ of equation (20) can have only bounded magnitude in practice as well. At the same time, the human command $(\dot{v}_2^d, \dot{w}_2^d)$ is in general arbitrary, thus, not necessarily always feasible (i.e. on the line of $\bar{u}_L^1$ in Figure 7). Due to these reasons, even if the exponential convergence of the human command tracking is theoretically established in Section 3.2.2, in practice, there always exists some lag and difference between the human command (v_2^d, w_2^d) and the real FPV camera motion (v_2, w_2).

This command-behavior mismatch may then confuse and even frustrate human users, particularly since it would be difficult for them to understand why the system does not track their command only with the limited FPV. This command-behavior mismatch would be even more exacerbated in practice due to, e.g. unpredicted disturbance, unmodeled dynamics, communication/actuator/sensing delays, etc. To address this here, we design some visual and haptic feedback, which supplements the FPV in such a way that the user can be replenished with more information for better understanding of this command-behavior mismatch, thereby, rendering the two WMRs more intuitive and efficient for them to teleoperate.

To design the visual and haptic feedback, we first map the feasible set (i.e. the line of $\bar{u}_L^1$) of Figure 7 into the master device space of Figure 6. It can then be shown that this mapped feasible set is given by the line spanned by the following unit vector

$$e_f(\phi_1 - \phi_2) := \begin{bmatrix} \frac{\eta_1 \, s(\phi_1 - \phi_2)}{\sqrt{\eta_2^1 + L^2 \eta_2^2}} & \frac{L\eta_2 \, c(\phi_1 - \phi_2)}{\sqrt{\eta_2^1 + L^2 \eta_2^2}} \end{bmatrix}^{\mathrm{T}} \in \Re^2$$

We also define the deviation of the user command (x_h, y_h) from this feasible set span$[e_f]$ to be

$$d := \begin{pmatrix} x_h \\ y_h \end{pmatrix} - e_f^{\mathrm{T}} \begin{pmatrix} x_h \\ y_h \end{pmatrix} e_f \in \Re^2 \tag{23}$$

see Figure 6.

We then design the haptic cue for the master interface s.t.

$$\tau_h = -k_h d \tag{24}$$

where $k_h > 0$ is the positive haptic gain. We also present the visual cue to the user of this feasible set span$[e_f]$ and the user command (x_h, y_h) as the line and the sphere respectively, along with the FPV camera view (see Figure 9). Note that the tracking control equation (20) attempts to draw the feasible set to the human command, whereas the haptic feedback equation (24) (and also implicitly the visual feedback) nudges the human command to the feasible set. This visuo-haptic feedback, by providing information on how far and in which direction the human command deviates from the feasible set, turns out to significantly enhance the user's awareness and understanding of the command-behavior mismatch as compared to the case of using only the (limited) FPV camera view, thereby, allowing for substantial performance and HRI experience improvement as experimentally validated in the next section.

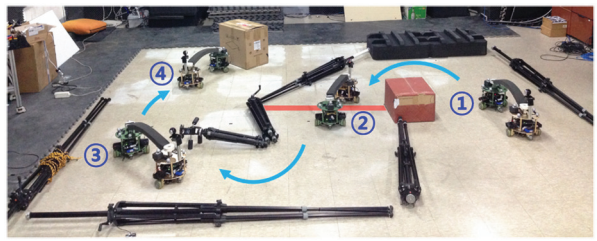

Fig. 8. Snapshots of the two cooperative nonholonomic WMRs squeeze-grasping/transporting a deformable object (sponge) with no grip-enforcing fixture during an experiment: for the human subject experiment of Section 4.3, the gain change from Λ_{HG} to Λ_{LG} occurs when the fore-running WMR crosses the line at ②.

4. Experiment and human subject study

In this section, we perform experiments to validate the proper functioning of our proposed FPV semi-autonomous teleoperation framework derived in Section 3 to allow a human user to tele-drive the deformable object via the FPV camera while maintaining the fixture-less grasping and FPV-centering. We also perform a human subject study to show that the visuo-haptic feedback of Section 3.2 can

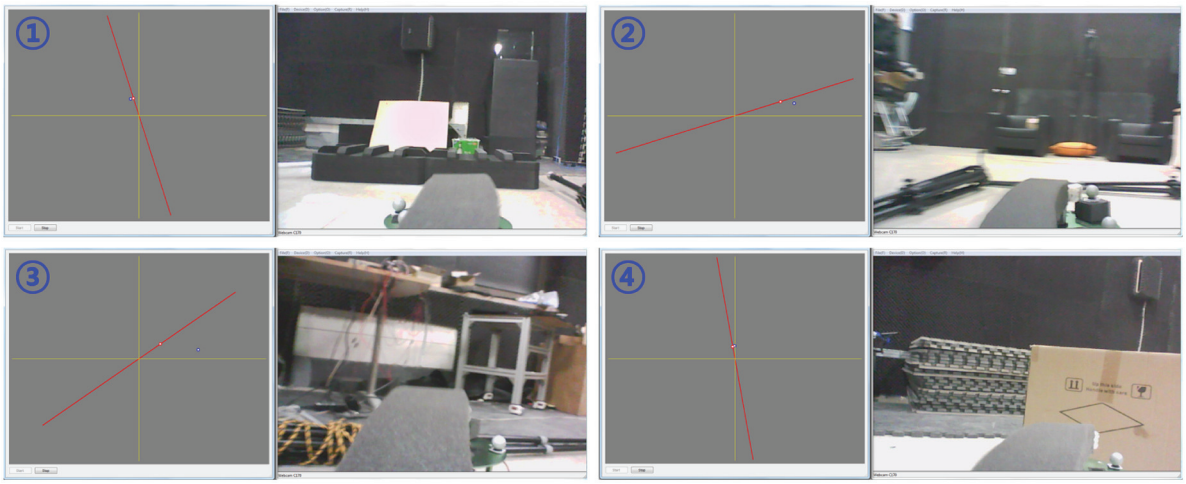

Fig. 9. Snapshots of the master user view, consisting of the FPV camera view, and the visual cue of the human command $(x_h, y_h) = (\frac{w_2^d}{\eta_2}, \frac{v_2^d}{\eta_1})$ and the feasible set span$[e_f]$ represented by the white sphere and the red-line (see Figure 6). Here, the circled numbers correspond to those in Figure 8. Note that the fore-running WMR is kept at the center of the FPV-FOV (i.e. FPV-centering) as claimed in Section 3.2.2.

significantly improve the performance of the FPV semi-autonomous teleoperation by enhancing the user perception of, and also nudging them to reduce, the command-behavior mismatch, which is otherwise difficult for them to understand/predict only with the limited FPV, thus, can cause their confusion/frustration and performance degradation. Another interesting finding from this human subject study is that, for our system, the visual feedback is as effective as the haptic feedback, suggesting that, for some teleoperation systems, one can just adopt visual feedback instead of haptic feedback, with the former typically cheaper to realize than the latter, although their combined visuo-haptic feedback would provide the best performance teleoperation system.

4.1. Experimental setup

We utilize the two differential-drive WMRs as shown in Figure 1 as the slave system, each possessing two passive front/rear caster wheels and two side active wheels driven by Maxon® BLDC motors and 18:1 gearboxes under the velocity control mode with the wheel velocity command in the PWM (pulse width modulation) form received from Arduino Uno MCU (micro-controller unit) for each WMR. The two WMRs are of similar size and squeeze-holding the deformable sponge together in the fixture-less grasping manner. A standard web-cam camera (Logitech C170, 640×480, 30FPS) is also attached on the over-seeing WMR as the FPV camera. The squeezing distance L is also set to be 57.5 cm.

We use a Force Dimension® Omega 3 haptic device as shown in Figure 2, which allows human users to command the FPV camera view while also providing haptic feedback. We only utilize the horizontal 2-DOF of the haptic device, and also match the forward/backward and the left/right directions of the haptic device (with respect to the human user) with the forward/backward linear motion and the counterclockwise/clockwise yaw motion of the FPV camera respectively, as explained in equation (15) and shown in Figure 6. When chosen, the haptic feedback equation (24) is provided with a rate of 1 kHz in the device space as shown in Figure 6, whereas, if the option of visual feedback is chosen, the human command (x_h, y_h) and the feasible set as defined in equation (21) are graphically rendered in the device space as shown in Figure 6 with a rate of 10 Hz. For all cases, the FPV camera view is shown to the human users at 30 Hz (see Figure 9 for the master view).

We also utilize the VICON® motion capture (MOCAP) system to measure the position/orientation of the WMRs with 240 Hz rate and sub-millimeter resolution. We also implement some simple estimation and low-pass filtering algorithms to address sudden occasional data missing from the MOCAP system. The control command for each WMR is then computed by the ground PC using the MOCAP data, which is then communicated to the Arduino Uno MCU of each WMR through XBee PRO with the baud rate of 57600 bit/s and a control refreshing rate of about 500 Hz.

4.2. Performance validation experiment

The goal of this experiment is to verify the proper functioning and performance of our proposed FPV semi-autonomous teleoperation framework, particularly in its capacity to rigidly maintain the fixture-less grasping and the FPV-centering under the human teleoperation and also ensure adequate tracking of the human command by the FPV camera under a mixed constraint. The efficacy of the visuo-haptic feedback is also alluded, which will be more

formally and concretely manifested in Section 4.3 through human subject study.

For this performance experiment, a human user is asked to tele-drive the two cooperative WMRs from point ① to point ④ along an s-like curved corridor while avoiding collisions against obstacles as shown in Figure 8. The following four cases are considered:

(a) only with the FPV camera view (i.e. NH/NV);
(b) the FPV camera with only the haptic feedback (i.e. H/NV);
(c) the FPV camera with only the visual feedback (i.e. NH/V);
(d) the FPV camera view with both the haptic and visual feedback activated (i.e. H/V).

Examples of the master interface view with the FPV camera view and the visual feedback (of the human command and the feasible set: not shown for the NH/NV and H/NV cases) are shown in Figure 9, where ①, ②, ③, ④ correspond to the same points in Figure 8. Example trajectories of the two WMRs during this teleoperation (for the case of NH/NV) are also shown in Figure 10.

We set the potential function φ_h for the grasping control equation (14) and the gain Λ for the human command tracking control equation (20) as best as we can, as the goal here is to experimentally show that the proposed architecture is adequately functioning with a proper gain setting. We also implement the haptic feedback gain k_h equation (24) to be $k_h = 52$ N/m, which is found to be most effective, yet at the same time, not intrusive to the human user. However, this optimal value of haptic gain k_h is subjective and will be chosen by each subject for the human subject study in Section 4.3. For each of the four cases, we allow the human user to spend as much preparatory time as he wants so that we can measure an "expert" user teleoperation behavior from the user, although, as shown in Section 4.3, this "expert" behavior in general significantly differs among different users. We then ask the user to commence the task and measure the data.

The results of the performance experiment are presented in Figures 11 to 16, with some important data also summarized in Table 1, from which we can first observe that the grasping distance error $|e_L| = |\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} - L| \leq 1$[cm] for all the cases during all operations (Figure 11). This level of the grasping error is enough for the two WMRs to maintain the squeezing of the deformable sponge, implying that the rigid grasping can be maintained even without any physical grip-enforcing fixture regardless of the human teleoperation command. This guarantee of fixture-less grasping regardless of human teleoperation, that comes from our usage of nonholonomic passive decomposition of Section 2.3, is important, since the dropping of the object may even be unsafe in practice, yet, as shown in Section 4.3, the way of teleoperation can exhibit significant variability among different users. Here, note also that we evaluate e_L

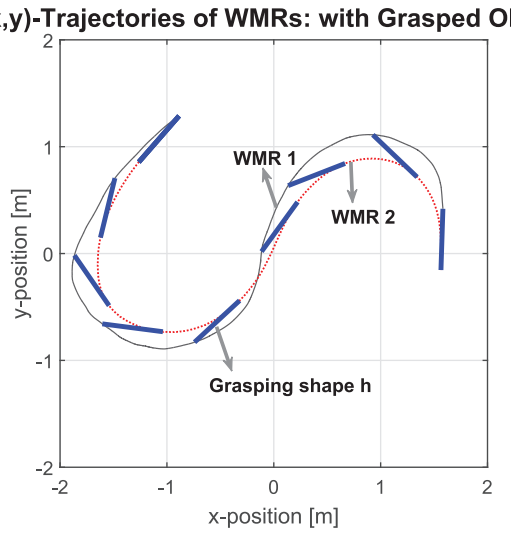

Fig. 10. Trajectories of the two WMRs during the performance experiment (NH/NV case): fixture-less grasping and FPV-centering are maintained during the operation as shown by the desired inter-robot grasping shape (blue sold lines). Other three cases also exhibited similar trajectories of the WMRs.

instead of $||h||$ in equation (4), as the grasping (or dropping) of the object is solely dictated by e_L with the adoption of rotating disk as shown in Figure 1. We can also see from the results that the FPV camera can track the human command fairly well, with the tracking errors $|e_v|, |e_w|$ kept small during the operations (Figure 12, Figure 13), and, consequently, the command-behavior mismatch $||d|| \leq 2$ cm most of the time throughout the experiments (Figure 14).

Even if it is kept small during all the experiments, we can still notice that the command-behavior mismatch $||d||$ is largest with no feedback (i.e. NH/NV) and the smallest with full-feedback (i.e. H/V) with those of the H/NV and NH/V cases being in between these (Figure 14). This in fact signifies the efficacy of the multi-modal feedback for the FPV teleoperation of the two WMRs under the mixed constraint, that is with no feedback it is very difficult for human users to understand/predict the command-behavior mismatch only with the limited FPV, thus they are prone to generate an infeasible command more often, which may cause more confusion and frustration as the system cannot track the command without their knowing why. On the other hand, with the visual or haptic feedback, they can see or feel how far and in which direction their command is from being feasible, thereby, it can be nudged to generate a more feasible command and the system can now better track their command and react to them in more anticipated manner, thereby, decreasing their confusion/frustration and increasing the system performance. The intensity of this trend is most significant with both the visual and haptic feedback together (i.e. H/V). This benefit of the visuo-haptic feedback will be more rigorously and comprehensively verified in the next section via human subject study.

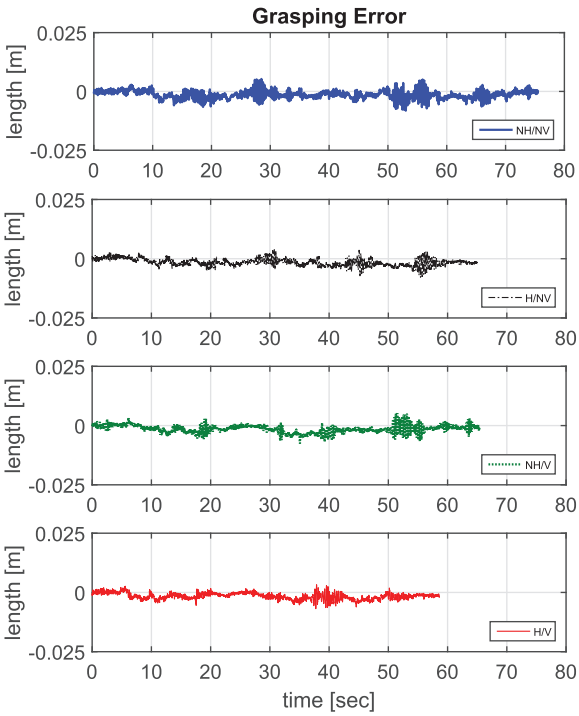

Fig. 11. Grasping distance error $e_L := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} - L$ for the performance experiment.

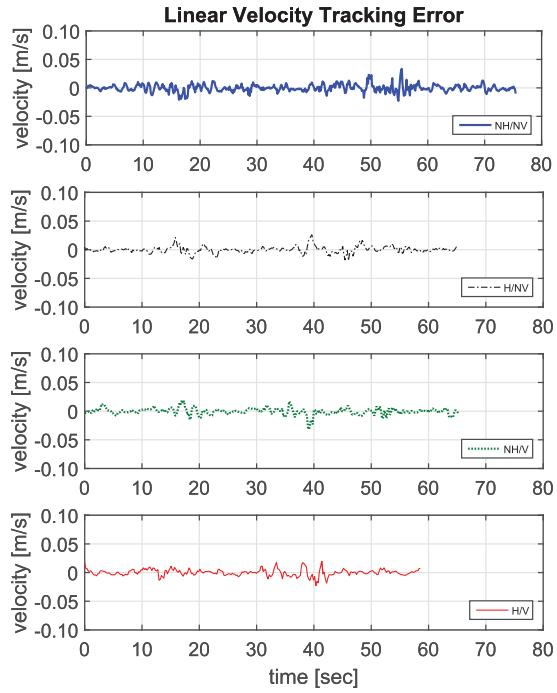

Fig. 12. Forward velocity human command tracking error e_v for the performance experiment.

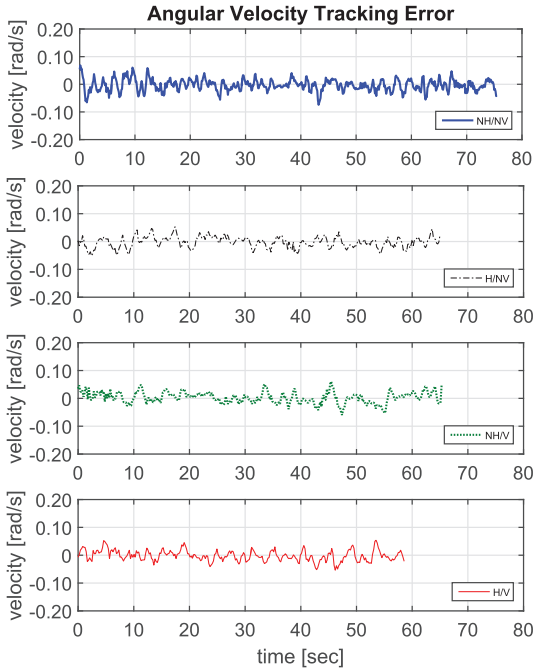

Fig. 13. Turning velocity human command tracking error e_w for the performance experiment.

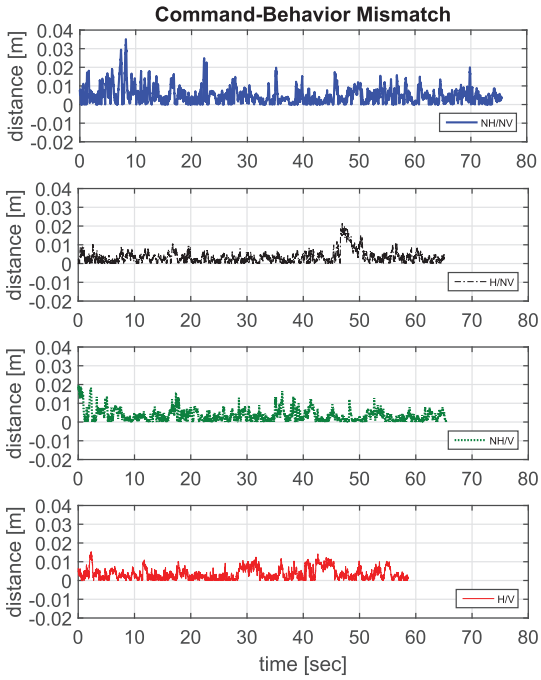

Fig. 14. Command-behavior mismatch $||d||$ for the performance experiment.

4.3. Human subject study: Efficacy of visuo-haptic feedback

In this section, we perform a human subject study to verify the efficacy of the visual and haptic feedback for the FPV semi-autonomous teleoperation of the WMRs under the mixed constraint. For this, we choose the same task as Section 4.2, i.e. to tele-drive through the s-like shaped corridor in Figure 8. We also use the same setup as in Section 4.2. The four cases of NH/NV, H/NV, NH/V and H/V are also considered in the human subject study. We recruit seven

Table 1. Performance experiment data: grasping distance error $e_L := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} - L$, human command tracking error $|e_v|$ and $|e_w|$ defined in equation (18) and the command-behavior mismatch $||d||$ defined in equation (23).

Criteria	NH/NV	H/NV	NH/V	H/V
$\|e_L\|_{\text{RMS}}$ (m)	0.002195	0.002161	0.002179	0.002078
$\|e_L\|_{\text{MAX}}$ (m)	0.008198	0.007985	0.007551	0.006855
$\|e_v\|_{\text{RMS}}$ (m/s)	0.006356	0.005814	0.005899	0.005763
$\|e_v\|_{\text{MAX}}$ (m/s)	0.033519	0.026255	0.031364	0.022857
$\|e_\omega\|_{\text{RMS}}$ (rad/s)	0.021435	0.019833	0.020651	0.019568
$\|e_\omega\|_{\text{MAX}}$ (rad/s)	0.074306	0.053766	0.060287	0.053194
$\|\|d\|\|_{\text{RMS}}$ (m)	0.006548	0.004647	0.004693	0.004570
$\|\|d\|\|_{\text{MAX}}$ (m)	0.035118	0.021458	0.019595	0.015293

human subjects, all aged between 20 and 39 with no known neural disorders or disability. None of them were involved in the research development of this paper. They can only oversee the operation via the FPV camera attached on the following WMR, while also wearing earplugs to mask the ambient and WMR-generating noises. We also conduct all of the human subject study experiments in accordance with the requirements of the Helsinki Declaration.

Each participant is then given as much time as they want to get familiar with the setup and the teleoperation task under each of the four cases. For this preparatory period, the same potential function φ_h for the grasping control equation (14) and the same gain matrix Λ for the human command tracking control equation (20) as for the performance experiment of Section 4.2 are used. Let us denote this high-performance gain by Λ_{HG} in the sequel. During this preparatory period, each subject also chooses their preferred haptic gain k_h for equation (24) when they practice with the cases of H/NV and H/V. For this, we restrict k_h to be larger than 42 N/m, since, below this value, the effect of haptic feedback is found to be too weak to be effective. This chosen single value of the haptic feedback gain k_h is then used for that subject both for the cases of H/NV and H/V. The preparatory period lasts for around 15 min typically for each subject.

After this preparatory period, each subject then proceeds to one of the four cases, presented in a random order. We only ask each subject to complete the task within 100 s and without any collisions. However, we observe that all the subjects tend to speed up the operation to the extent comfortable to avoid the collisions without intentionally wandering around. We consider the collisions with obstacles and a completion time longer than 100 s as a task failure. When the subject fails, we give them a chance to re-do the same experiment, with the failed attempt considered as a part of the preparatory practice. This re-doing of the experiment does not interfere with the conclusion of our human subject study, since even if it unfairly gives an advantage

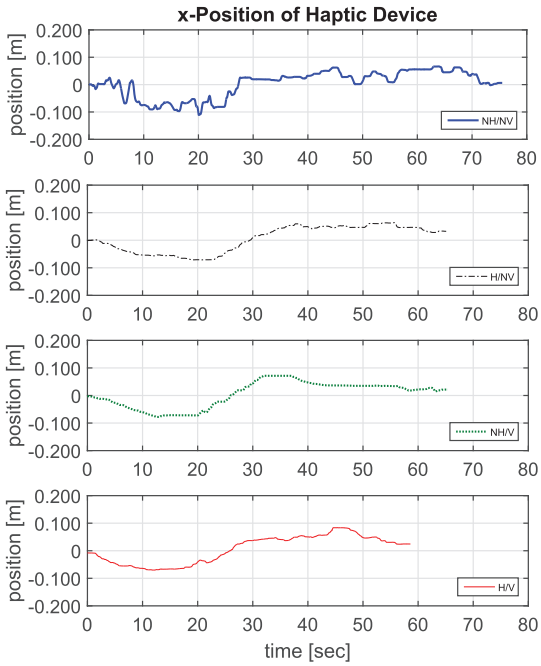

Fig. 15. Human user turning velocity command x_h during the performance experiment.

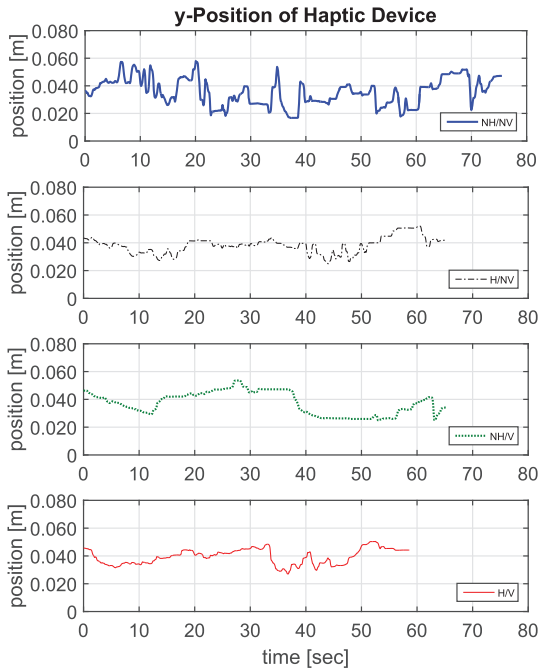

Fig. 16. Human user forward velocity command y_h during the performance experiment.

to the case of NH/NV (i.e. more learning with higher failure rates), the case of NH/NV still exhibits the highest rate of failure with the worst subjective rating among the four cases (see Figure 23 and Figure 24). The consecutive five failures are however defined as an ultimate failure and the subject proceeds to the next case.

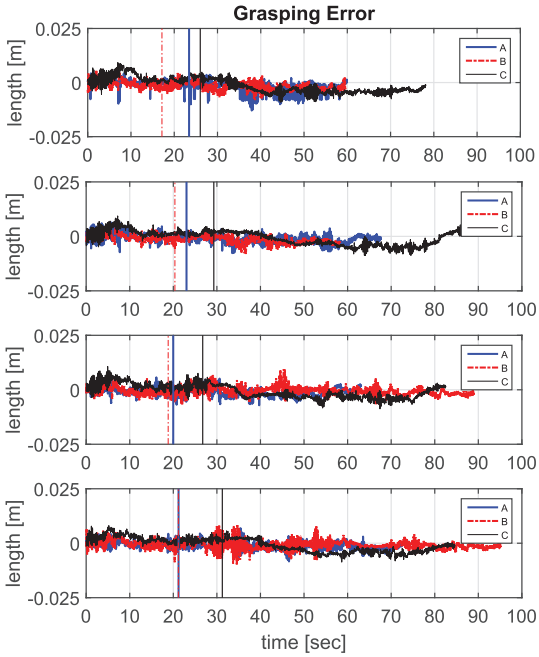

Fig. 17. Grasping distance error $e_L := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} - L$ of the subjects A, B and C with vertical lines indicating instants of gain degradation (NH/NV, H/NV, NH/V, H/V from the top).

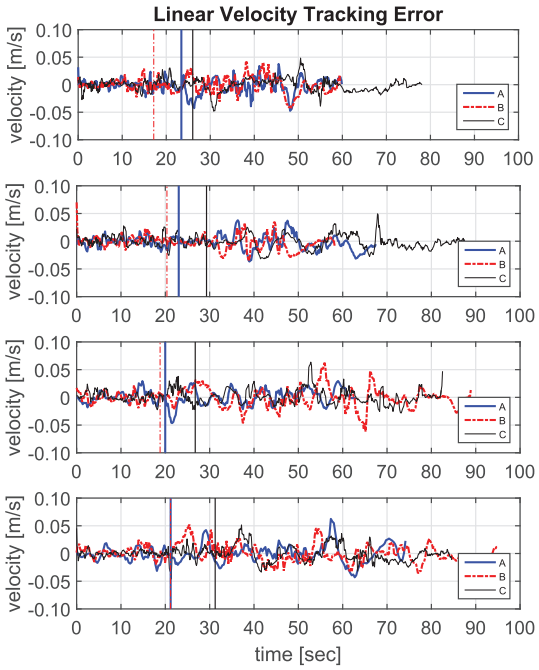

Fig. 18. Forward velocity human command tracking error e_v of the subjects A, B and C with vertical lines indicating instants of gain degradation (NH/NV, H/NV, NH/V, H/V from the top).

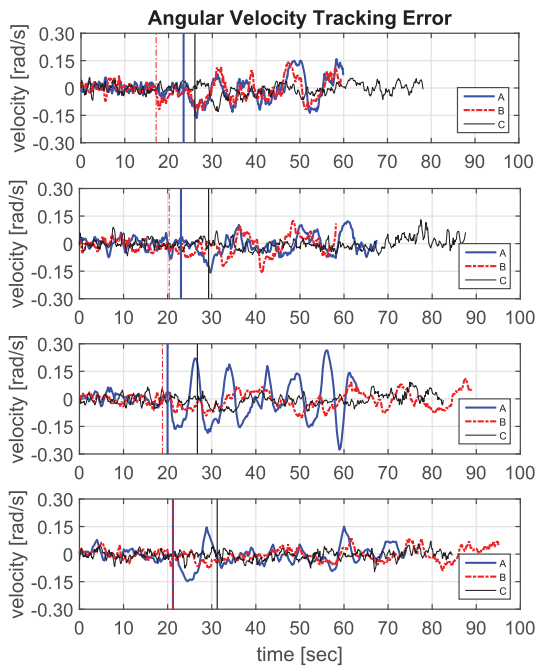

Fig. 19. Turning velocity human command tracking error e_w of the subjects A, B and C with vertical lines indicating instants of gain degradation (NH/NV, H/NV, NH/V, H/V from the top).

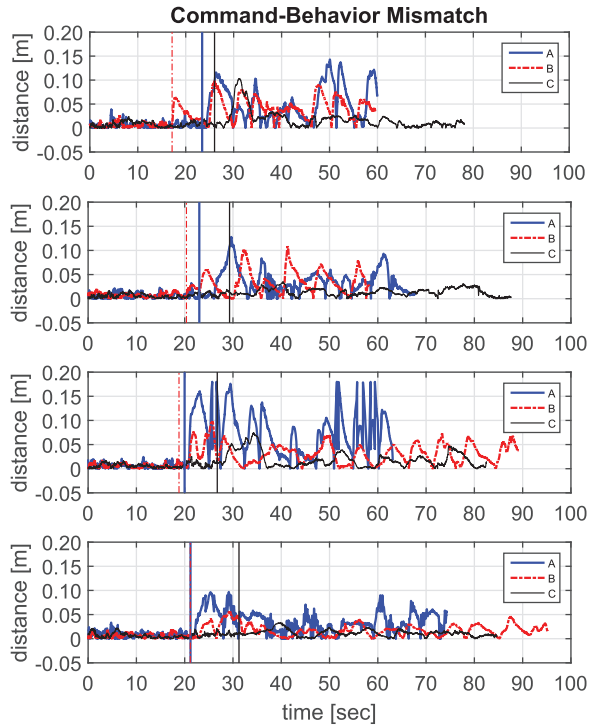

Fig. 20. Command-behavior mismatch $||d||$ of the subjects A, B and C with vertical lines indicating instants of gain degradation (NH/NV, H/NV, NH/V, H/V from the top).

For each case of the human subject experiment, when the fore-running WMR hits the red line at point ② of Figure 8, we degrade the human command tracking gain from Λ_{HG} to Λ_{LG} exponentially within 1.5 s, with the feedforward action (i.e. $(\dot{v}_2^d; \dot{w}_2^d)$) of equation (20) also deactivated

at the end of this time window. We intentionally degrade the human command tracking performance in this way to emulate such frequent non-idealities in practical teleoperation scenarios in unknown/unmodeled/dynamic environments as unforeseen disturbance (e.g. excessive wheel slips, wear of driving shafts), unmodeled dynamics (e.g. unpredicted slops, sudden increase of payload), information delay (e.g. network latency, actuator/sensing delays), etc. We then evaluate/analyze the data of only this segment with the degraded human command tracking with Λ_{LG}.

After each trial, it is also requested that the subject answers the NASA-TLX questionnaire (Hart and Staveland, 1988; Rubio and Puente, 2004). For this, the subject is asked to evaluate only for the segment with the degraded human command tracking Λ_{LG}. The six standard items of the NASA-LTX are used (see Figure 23):

- mental demand;
- physical demand;
- temporal demand;
- performance;
- effort
- frustration.

For the performance index, we define the score to be 20 times the number of failures for one subject in each case, with the score of the ultimate failure case (i.e. five consecutive fails) being 100. After completing all four cases, each subject is also given a survey, which is similar to the NASA-TLX questionnaire, regarding how they are satisfied with the user interface and how helpful it is for the four cases (see the bottom plot of Figure 23). For more details on the NASA-TLX test, we refer readers to the work by Hart and Staveland (1988) and Rubio and Puente (2004).

The results of the human subject study are then shown in Figures 17 to 22 (detailed data plots of the three representative subjects); Table 2 (summary for some key performance measures); Figures 23 to 24 (NASA-TLX and survey results) and Figures 25 to 26 (data of the six human subjects). First, from Table 2 and Figure 25, we can notice that the maximum grasping distance error e_L is still maintained to be less than 1.5 cm, implying that the fixture-less grasping can still be guaranteed even in the presence of the human command tracking degradation (see Figures 18 and 19) and the significant variability of the human teleoperation commands (see Figures 21 and 22). Again, this guarantee of the fixture-less grasping stems from our adoption of the behavior decomposition approach (Section 2.3) and would be important for real teleoperation scenarios, where the behaviors of different users can differ substantially from each other as observed in Figures 21 and 22.

We can also clearly see from Figures 18 and 19 that, once the intentional gain degradation with Λ_{LG} kicks in, the error (e_v, e_w) of the human command tracking increases, and so does the command-behavior mismatch $||d||$ as shown in Figure 20. This can also be seen by comparing Table 2 (or Figure 25) with the Table 1. Similarly, we can also observe

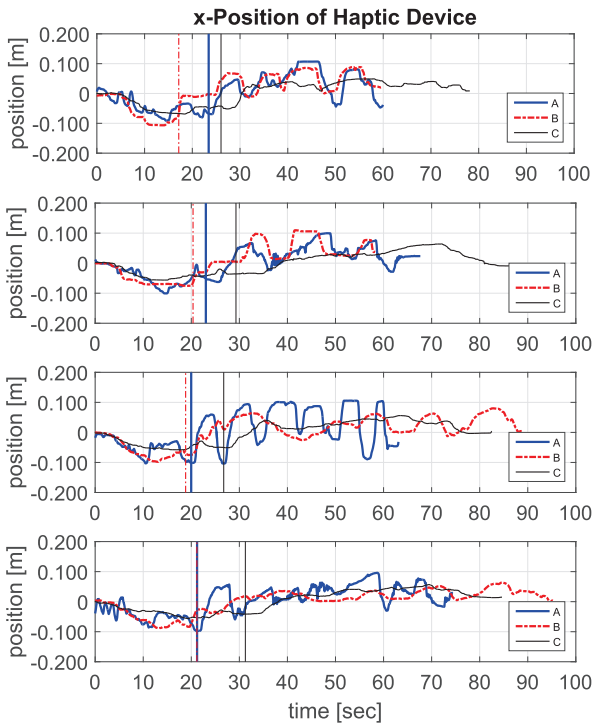

Fig. 21. Human user turning velocity command x_h of the subjects A, B and C with vertical lines indicating instants of gain degradation (NH/NV, H/NV, NH/V, H/V from the top).

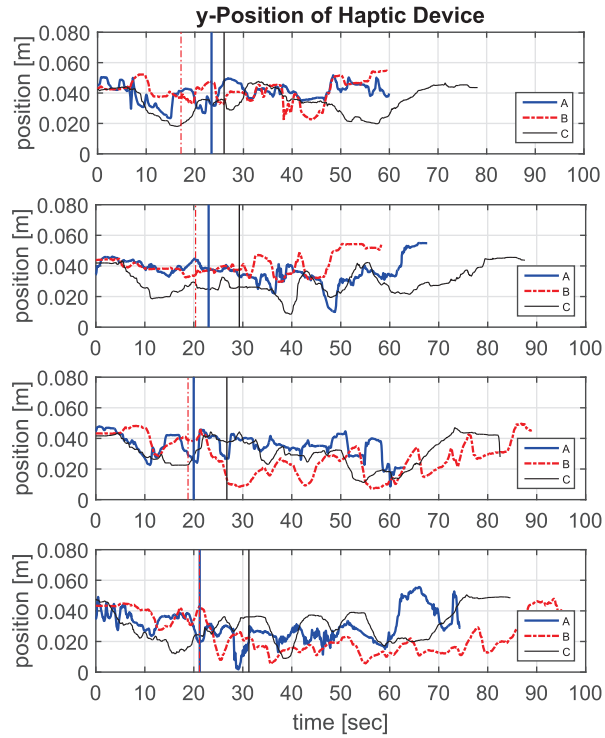

Fig. 22. Human user forward velocity command y_h of the subjects A, B and C with vertical lines indicating instants of gain degradation (NH/NV, H/NV, NH/V, H/V from the top).

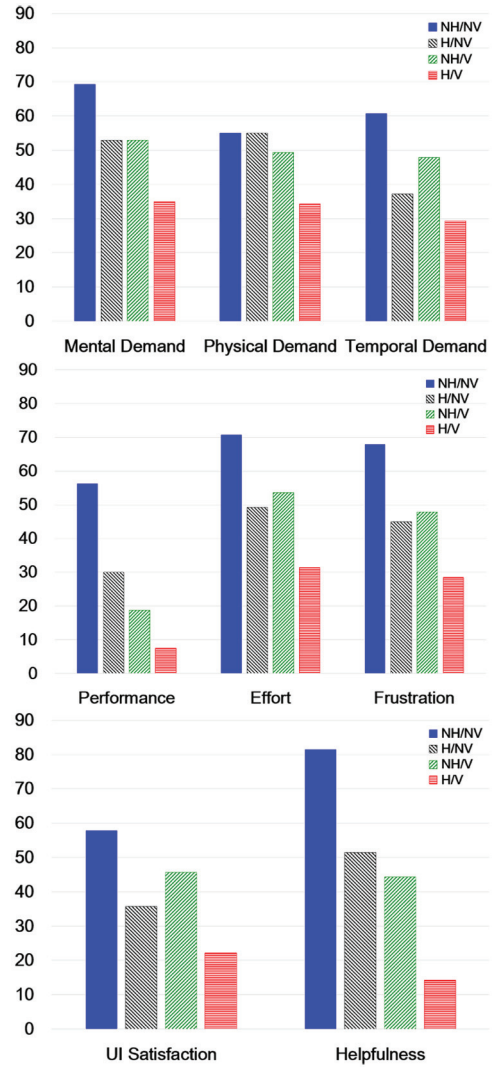

Fig. 23. Evaluation of the four cases based on the NASA-TLX questionnaire (top and middle); and survey results on user interface satisfaction and helpfulness of the four cases (bottom). For all cases, the smaller, the better.

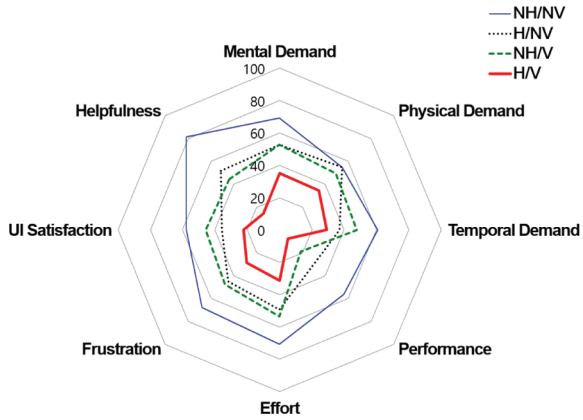

Fig. 24. Polygon plot of the four cases based on the eight items in Figure 23: the smaller the polygon, the better the performance it indicates.

Table 2. Human subject study experiment data: grasping distance error $e_L = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} - L$, human command tracking error $|e_v|$ and $|e_w|$ (defined in equation (18)), the command-behavior mismatch $||d||$ (defined in equation (23)) and the task completion time T. Here, $|e_L|^{\text{AVE}}_{\text{MAX}}$, $|e_v|^{\text{AVE}}_{\text{MAX}}$, $|e_\omega|^{\text{AVE}}_{\text{MAX}}$ and $||d||^{\text{AVE}}_{\text{MAX}}$ are the average of $|e_L|_{\text{MAX}}$, $|e_v|_{\text{MAX}}$, $|e_\omega|_{\text{MAX}}$ and $||d||_{\text{MAX}}$ across all the human subjects, respectively.

Criteria	NH/NV	H/NV	NH/V	H/V
$\|e_L\|_{\text{RMS}}$ (m)	0.003162	0.003011	0.002774	0.002682
$\|e_L\|^{\text{AVE}}_{\text{MAX}}$ (m)	0.010270	0.010121	0.009676	0.008943
$\|e_v\|_{\text{RMS}}$ (m/s)	0.018598	0.017591	0.016786	0.016511
$\|e_v\|^{\text{AVE}}_{\text{MAX}}$ (m/s)	0.063864	0.060646	0.050353	0.045046
$\|e_\omega\|_{\text{RMS}}$ (rad/s)	0.062530	0.057567	0.058462	0.051390
$\|e_\omega\|^{\text{AVE}}_{\text{MAX}}$ [rad/s]	0.162902	0.149765	0.163775	0.143228
$\|\|d\|\|_{\text{RMS}}$ (m)	0.049297	0.041817	0.046364	0.037677
$\|\|d\|\|^{\text{AVE}}_{\text{MAX}}$ (m)	0.134272	0.107576	0.126546	0.105373
T_{AVG} (s)	75.47	71.05	79.37	73.98
T_{MIN} (s)	59.54	58.26	63.09	58.88
T_{MAX} (s)	98.21	87.59	89.35	95.18
T_{STDV} (s)	13.11	8.12	8.87	12.45

that the command-behavior mismatch $||d||$ increases with this degraded human command tracking with Λ_{LG} from Table 2 (or Figure 25) as compared to Table 1.

On the other hand, the decreasing of the command-behavior mismatch $||d||$ with the feedback can still be noticed from Table 2 (i.e. $||d||^{\text{AVE}}_{\text{MAX}}$ is maximum for the case of NH/NV, minimum for H/V, with those for NH/V, H/HV in between), yet not as vivid as for the performance experiment (Table 1, see also Figure 25). This we believe is because, for the performance experiment, we only have one human user, who happens to exhibit the aforementioned trend vividly. However, for the human subject study, there exists significant variability on how the six users behave during the teleoperation (e.g. see Figures 21 and 22), with some users showing similar trends (e.g. smooth command while using the visual/haptic feedback as guidance) while others not (e.g. aggressive command while using the visual/haptic feedback to correct their command only when too far from the feasible set). Therefore, the trends are rather averaged out across all the different users.

Other than this, we cannot find further dominant trends among the four cases (i.e. NH/NV, H/NV, NH/V and H/V) from Table 2, Figures 17 to 22 and Figures 25 and 26. However, we can observe unequivocal differences and trends among these four cases from the NASA-TLX and survey questionnaire tests (see Figures 23 and 24), from which we can see the obvious preference for the full visuo-haptic feedback (i.e. H/V) over the case of only the FPV camera

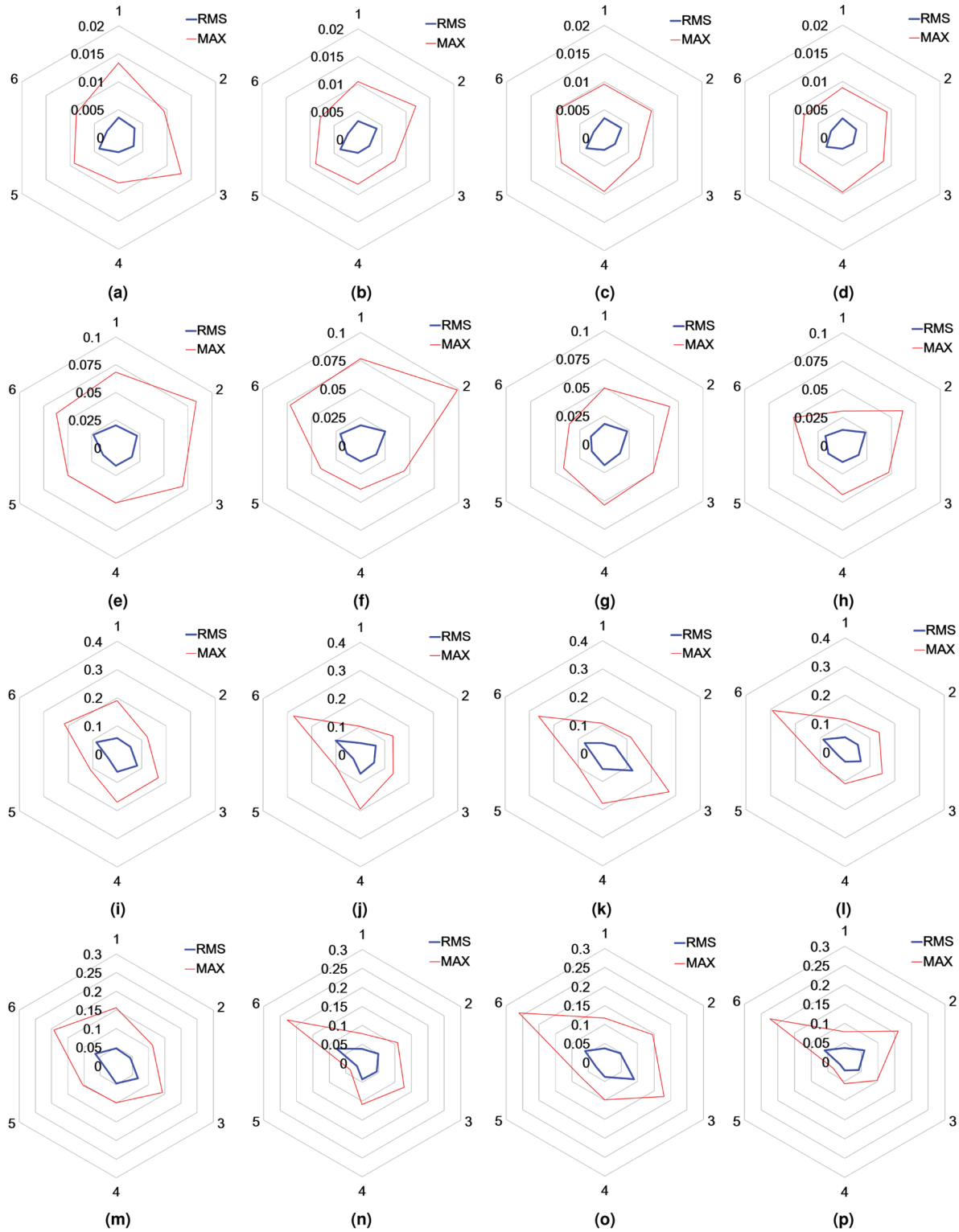

Fig. 25. Grasping distance error e_L, human command tracking error (e_v, e_ω) and the command-behavior mismatch $||d||$ of the six human subjects. (a) e_L (NH/NV), (b) e_L (H/NV), (c) e_L (NH/V), (d) e_L (H/V), (e) e_{v_2} (NH/NV), (f) e_{v_2} (H/NV), (g) e_{v_2} (NH/V), (h) e_{v_2} (H/V), (i) e_{ω_2} (NH/NV), (j) e_{ω_2} (H/NV), (k e_{ω_2} (NH/V), (l) e_{ω_2} (H/V), (m) $|d|$ (NH/NV), (n) $|d|$ (H/NV), (o) $|d|$ (NH/V) and (p) $|d|$ (H/V).

view (i.e. NH/NV). It is also very interesting that only the visual feedback (i.e. NH/V) and only the haptic feedback (i.e. H/NV) receive more or less a similar preference rating (see Figure 24). This latter finding, we believe, has an

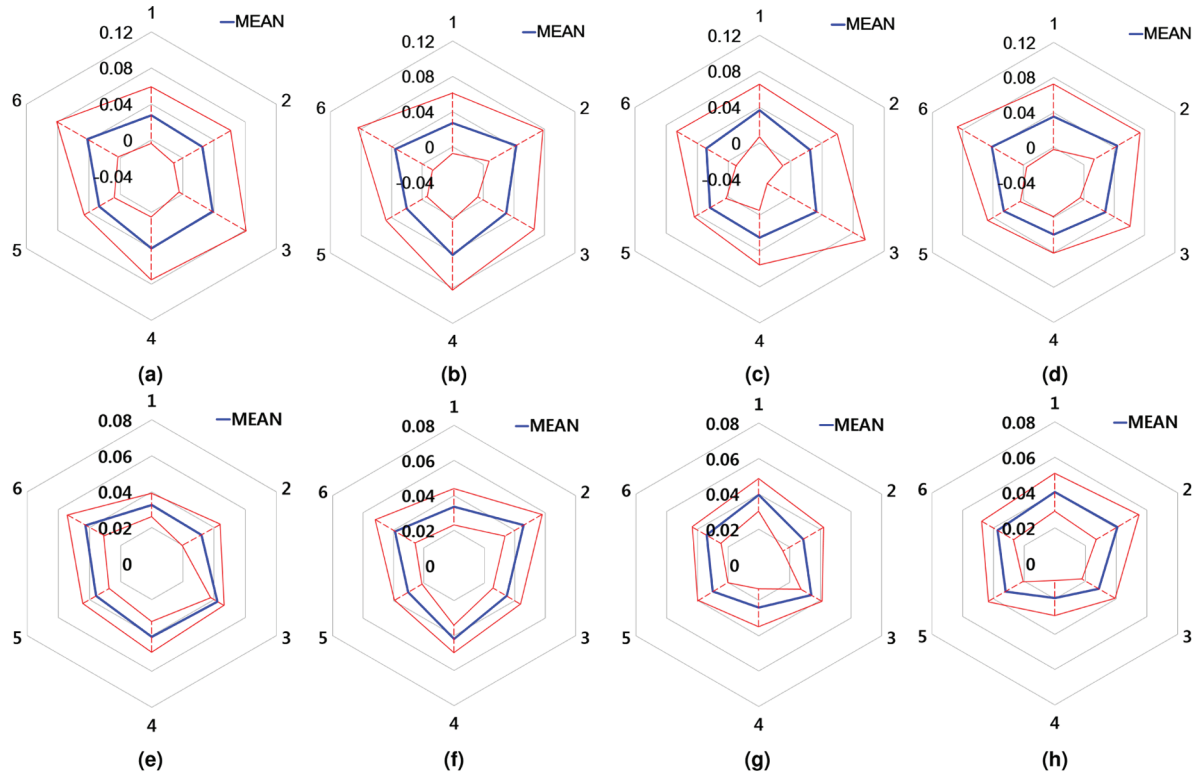

Fig. 26. Human teleoperation command x_h and y_h of the six human subjects. (a) x_h (NH/NV), (b) x_h (H/NV), (c) x_h (NH/V), (d) x_h (H/V), (e) y_h (NH/NV), (f) y_h (H/NV), (g) y_h (NH/V) and (h) y_h (H/V).

important practical ramification, that is, for some teleoperation systems, the visual feedback, which is typically cheap to generate using the currently-available display and graphics rendering technique, may provide performance and user preference comparable to those with haptic feedback, which is still relatively expensive to realize due to its necessitating actuator hardwares.

This effectiveness of the questionnaire-based evaluation we believe is because, even if all the subjects exhibit such a wide-variability in their behaviors during the teleoperation from each other so that some meaningful trends cannot be easily extracted from their motion-related data by simple scalar numbers (e.g. Figures 21 and 22), they still perceive some differences among the cases, for which the verbal questionnaire, that is designed to better capture such perception related issues (Hart and Staveland, 1988; Rubio and Puente, 2004), is more effective. Further human subject studies for different system configurations (e.g. multiple WMRs), master interface (e.g. AR, augmented reality, visual cue) or scenarios (e.g. transporting with middle-point examining) using more quantitative/numeric approaches (e.g. psychophysics approaches Jang and Lee, 2014; Lee et al., 2015) are also possible and a topic for future research.

5. Summary and future research

We propose a novel FPV semi-autonomous teleoperation framework for two nonholonomic WMRs to cooperatively and fixture-lessly grasp/transport a deformable object while also being teleoperated by a remote user via a FPV camera, which is attached on the over-seeing WMR. These two WMRs then turn out to be under the mixed constraint, i.e. physical/nonholonomic constraint (i.e. no-slip/drift condition) and artificial/holonomic requirement (i.e. grasping/FPV-centering). To address this mixed constraint, nonholonomic passive decomposition (Lee 2010a,c) is utilized to achieve behavior decomposition of the two WMRs into grasping-related behavior and teleoperation-related behavior, and propose semi-autonomous teleoperation architecture based on that. Certain visuo-haptic feedback is also designed to notify the users of the command-behavior mismatch, which inevitably arises due to the mixed constraint, but is difficult to understand only with the limited FPV. Thus, this can significantly degrade intuitiveness and performance of FPV teleoperation. Human subject study is performed to manifest the efficacy of the visuo-haptic feedback in addressing this command-behavior mismatch.

Some possible future research topics include:

(a) on-board implementation and outdoor experimentation;
(b) extension to multi-WMRs with multiple cameras;
(c) application/extension to other robotic systems under mixed constraint;
(d) incorporation of user personality into teleoperation control design.

Authors' note

Author Kent Yee Lui and Hyunjun Cho are now affiliated with ASML Korea, Republic of Korea and Hanwha Cooperation, Republic of Korea.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Basic Science Research Program (grant number 2015R1A2A1A15055616), the Engineering Research Center Program for Soft Robotics (grant number 2016R1A5A1938472) of the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT & Future Planning (MSIP) and the Industrial Strategic Technology Development Program (grant number 10060070) of the Ministry of Trade, Industry & Energy (MOTIE) of Korea.

References

Alonso-Mora J, Knepper R, Siegwart R, et al. (2015) Local motion planning for collaborative multi-robot manipulation of deformable objects. In: *Proceedings of the IEEE international conference on robotics & automation*, Seattle, WA, USA, 26–30 May 2015, pp.5495–5502.

Anderson RJ and Spong MW (1989) Bilateral control of teleoperators with time delay. *IEEE Transactions on Automatic Control* 34(5): 494–501.

Buss M, Peer A, Schaub T, et al. (2010) Development of a multimodal multi-user telepresence and teleaction system *International Journal of Robotics Research* 29(10): 1298–1316.

Franchi A, Secchi C, Ryll M, et al. (2012) Shared control: Balancing autonomy and human assistance with a group of quadrotor UAVs. *IEEE Robotics & Automation Magazine* 19(3): 57–68.

Franchi A, Secchi C, Son HI, et al. (2012) Bilateral teleoperation of groups of mobile robots with time-varying topology. *IEEE Transactions on Robotics* 28(5): 1019–1033.

Goertz R (1952) Fundamentals of general-purpose remote manipulators. *Nucleonics* 10(11): 36–42.

Goertz R and Bevilacqua F (1952) A force-reflecting positional servomechanism. *Nucleonics* 10(11): 43–45.

Ha C and Lee DJ (2013) Vision-based teleoperation of unmanned aerial and ground vehicles. In: *Proceedings of the IEEE international conference on robotics & automation*, Karlsruhe, Germany, 6–10 May 2013, pp.1457–1462.

Ha C, Zuo Z, Choi FB, et al. (2014) Passivity-based adaptive backstepping control of quadrotor-type UAVs. *Robotics & Autonomous Systems* 62: 1305–1315.

Hart SG and Staveland LE (1988) Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in Psychology* 52: 139–183.

Hirata Y, Kume Y, Wang ZD, et al. (2005) Handling of a single object by multiple mobile manipulators in cooperation with human based on virtual 3-D caster dynamics. *JSME International Journal – Series C* 48(4): 613–619.

Jang I and Lee DJ (2014) On utilizing pseudo-haptics for cutaneous fingertip haptic device. In: *Proceedings of the haptics symposium*, Houston, TX, USA, 23–26 February, pp.635–639.

Karras GC, Loizou SG and Kyriakopoulos KJ (2011) Towards semi-autonomous operation of under-actuated underwater vehicles: Sensor fusion, on-line identification and visual servo control. *Autonomous Robots* 31: 67–86.

Khademian B and Hashtrudi-Zaad K (2011) Shared control architectures for haptic training: Performance and coupled stability analysis. *International Journal of Robotics Research* 30(13): 1627–1642.

Khatib O, Yokoi K, Chang K, et al. (1996) Coordination and decentralized cooperation of multiple mobile manipulators. *Journal of Robotic Systems* 13(11): 755–764.

Kim WS, Hannaford B and Bejczy AK (1992) Force-reflection and shared compliant control in operating telemanipulators with time delay. *IEEE Transactions on Robotics and Automation* 8(2): 176–185.

Kim Y and Minor MA (2010) Coordinated kinematic control of compliantly coupled multirobot systems in an array form. *IEEE Transactions on Robotics* 26(1): 173–180.

Kolmanovsky I and McClamroch NH (1995) Developments in nonholonomic control problems. *IEEE Control Systems Magazine* 15(6): 20–36.

Lawrence DA (1993) Stability and transparency in bilateral teleoperation. *IEEE Transactions on Robotics and Automation* 9(5): 624–637.

Lee DJ (2008) Semi-autonomous teleoperation of multiple wheeled mobile robots over the internet. In: *Proceedings of the ASME dynamic systems & control conference*, Ann Arbor, MI, USA, 20–22 October 2008, pp.147–154.

Lee DJ (2010a) Nonholonomic passive decomposition: Weak decomposability, controllability and control design. In: *Proceedings of the IEEE conference on decision & control*, Atlanta, GA, USA, 15–17 December 2010, pp.7123–7128.

Lee DJ (2010b) Passive configuration decomposition and practical stabilization of nonholonomic mechanical systems with symmetry. In: *Proceedings of the IEEE conference on decision & control*, Atlanta, GA, USA, 15–17 December 2010, pp.3620–3625.

Lee DJ (2010c) Passive decomposition and control of nonholonomic mechanical systems. *IEEE Transactions on Robotics* 26(6): 978–992.

Lee DJ and Li PY (2005) Passive bilateral control and tool dynamics rendering for nonlinear mechanical teleoperators. *IEEE Transactions on Robotics* 21(5): 936–951.

Lee DJ and Li PY (2013) Passive decomposition of mechanical systems with coordination requirement. *IEEE Transactions on Automatic Control* 58(1): 230–235.

Lee DJ and Spong MW (2005) Bilateral teleoperation of multiple cooperative robots over delayed communication networks: Theory. In: *Proceedings of the IEEE international conference on robotics & automation*, Barcelona, Spain, 18–22 April, pp.362–367.

Lee DJ and Spong MW (2006) Passive bilateral teleoperation with constant time delay. *IEEE Transactions on Robotics* 22(2): 269–281.

Lee DJ and Xu D (2011) Feedback *r*-passivity of Lagrangian systems for mobile robot teleoperation. In: *Proceedings of the IEEE international conference on robotics and automation*, Shanghai, China, 9–13 May, pp.2118–2123.

Lee DJ, Franchi A, Son HI, et al. (2013) Semi-autonomous haptic teleoperation control architecture of multiple unmanned aerial vehicles. *IEEE/ASME Transactions on Mechatronics* 18(4): 1334–1345.

Lee DJ, Martinez-Palafox O and Spong MW (2006) Bilateral teleoperation of a wheeled mobile robot over delayed communication networks. In: *Proceedings of the IEEE international conference on robotics & automation*, Orlando, FL, USA, 15–19 May, pp.3298–3303.

Lee Y, Jang I and Lee DJ (2015) Enlarging just noticeable differences of visual-proprioceptive conflict in VR using haptic feedback. In: *World haptics conference*, Evanston, IL, USA, 22–26 June 2015, pp.19–24.

Malysz P and Sirouspour S (2011a) A kinematic control framework for single-slave asymmetric teleoperation systems. *IEEE Transactions on Robotics* 27(5): 901–917.

Malysz P and Sirouspour S (2011b) Trilateral teleoperation control of kinematically redundant robotic manipulators. *International Journal of Robotics Research* 30(13): 1643–1664.

Murray RM, Li Z and Sastry SS (1993) *A Mathematical Introduction to Robotic Manipulation*. Boca Ranton: CRC.

Nielsen CW, Goodrich MA and Ricks RW (2007) Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics* 23(5): 927–941.

Pacchierotti C, Meli L, Chinello F, et al. (2015) Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems. *International Journal of Robotics Research* 34(13): 1773–1787.

Rubio S, Díaz E and Puente JM (2004) Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Advances in Psychology* 53(1): 61–86.

Sepulchre R, Jankovic M and Kokotovic P (1997) *Constructive Nonlinear Control*. London: Springer-Verlag.

Son HI, Franchi A, Chuang LL, et al. (2013), Human-centered design and evaluation of haptic cueing for teleoperation of multiple mobile robots. *IEEE Transactions on Cybernetics* 43(2): 597–609.

Sugar TG and Kumar V (2002) Control of cooperating mobile manipulators. *IEEE Transactions on Robotics and Automation* 18(1): 94–103.

Tang CP, Bhatt RM, Abou-Samah M, et al. (2006) Screw-theoretic analysis framework for cooperative payload transport by mobile manipulator collectives. *IEEE/ASME Transactions on Mechatronics* 11(2): 169–178.

Appendix

Index to multimedia extensions

Archives of IJRR multimedia extensions published prior to 2014 can be found at http://www.ijrr.org, after 2014 all videos are available on the IJRR YouTube channel at http://www.youtube.com/user/ijrrmultimedia

Table of multimedia extensions.

Extension	Media Type	Description
1	Video	Two nonholonomic WMRs under the FPV semi-autonomous teleoperation cooperatively holding/transporting a deformable object with no grip-enforcing fixture.

Lee DJ and Xu D (2011) Feedback *r*-passivity of Lagrangian systems for mobile robot teleoperation. In: *Proceedings of the IEEE international conference on robotics and automation*, Shanghai, China, 9–13 May, pp.2118–2123.

Lee DJ, Franchi A, Son HI, et al. (2013) Semi-autonomous haptic teleoperation control architecture of multiple unmanned aerial vehicles. *IEEE/ASME Transactions on Mechatronics* 18(4): 1334–1345.

Lee DJ, Martinez-Palafox O and Spong MW (2006) Bilateral teleoperation of a wheeled mobile robot over delayed communication networks. In: *Proceedings of the IEEE international conference on robotics & automation*, Orlando, FL, USA, 15–19 May, pp.3298–3303.

Lee Y, Jang I and Lee DJ (2015) Enlarging just noticeable differences of visual-proprioceptive conflict in VR using haptic feedback. In: *World haptics conference*, Evanston, IL, USA, 22–26 June 2015, pp.19–24.

Malysz P and Sirouspour S (2011a) A kinematic control framework for single-slave asymmetric teleoperation systems. *IEEE Transactions on Robotics* 27(5): 901–917.

Malysz P and Sirouspour S (2011b) Trilateral teleoperation control of kinematically redundant robotic manipulators. *International Journal of Robotics Research* 30(13): 1643–1664.

Murray RM, Li Z and Sastry SS (1993) *A Mathematical Introduction to Robotic Manipulation*. Boca Ranton: CRC.

Nielsen CW, Goodrich MA and Ricks RW (2007) Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics* 23(5): 927–941.

Pacchierotti C, Meli L, Chinello F, et al. (2015) Cutaneous haptic feedback to ensure the stability of robotic teleoperation systems. *International Journal of Robotics Research* 34(13): 1773–1787.

Rubio S, Díaz E and Puente JM (2004) Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Advances in Psychology* 53(1): 61–86.

Sepulchre R, Jankovic M and Kokotovic P (1997) *Constructive Nonlinear Control*. London: Springer-Verlag.

Son HI, Franchi A, Chuang LL, et al. (2013), Human-centered design and evaluation of haptic cueing for teleoperation of multiple mobile robots. *IEEE Transactions on Cybernetics* 43(2): 597–609.

Sugar TG and Kumar V (2002) Control of cooperating mobile manipulators. *IEEE Transactions on Robotics and Automation* 18(1): 94–103.

Tang CP, Bhatt RM, Abou-Samah M, et al. (2006) Screw-theoretic analysis framework for cooperative payload transport by mobile manipulator collectives. *IEEE/ASME Transactions on Mechatronics* 11(2): 169–178.

Appendix

Index to multimedia extensions

Archives of IJRR multimedia extensions published prior to 2014 can be found at http://www.ijrr.org, after 2014 all videos are available on the IJRR YouTube channel at http://www.youtube.com/user/ijrrmultimedia

Table of multimedia extensions.

Extension	Media Type	Description
1	Video	Two nonholonomic WMRs under the FPV semi-autonomous teleoperation cooperatively holding/transporting a deformable object with no grip-enforcing fixture.