A Teleoperating Interface for Ground Vehicles using Autonomous Flying Cameras

Daniel Saakes* Vipul Choudhary* Daisuke Sakamoto*† Masahiko Inami*‡ Takeo Igarashi*†

*JST ERATO Igarashi Design Interface Project ‡The University of Tokyo †Keio University

Abstract

Navigating remote robots and providing awareness of the remote environment is essential in many teleoperated tasks. An external view on the remote robot, a bird's eye view, is thought to improve operator performance. In this paper we explore a novel design for providing such a third-person view for a ground vehicle using a dynamic, external camera mounted on a quadcopter. Compared to earlier methods that use 3D reconstruction to create third-person views, our approach comprises a true third-person view through a video feed. We so provide visually rich, live information to the operator. In an experiment simulating a search and rescue mission in a simplified environment, we compared our proposed design to a pole-mounted camera and to a traditional front-mounted camera. The third-person perspective provided by our flying camera and pole-mounted camera resulted in fewer collisions and more victims being located, compared to the front-mounted camera.

Index Terms: H.1.2 [Information Systems]: Models and Principles—User/Machine systems; H.5.2 [Information Systems]: HCI—User Interfaces I.2.9 [Computing Methodologies]: Artificial Intelligence—Robotics

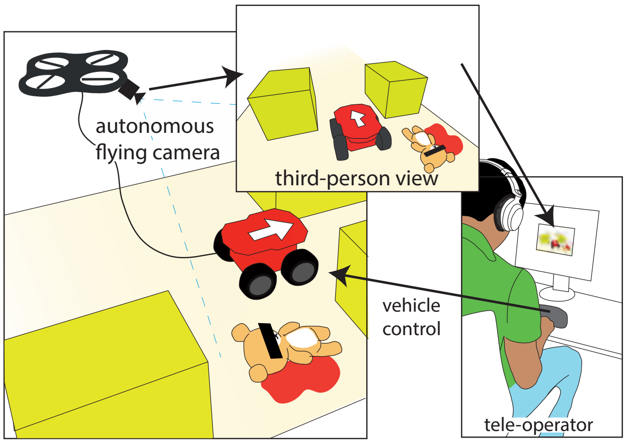

Figure 1: System overview. A teleoperator (right) controls a remote ground vehicle with a joystick. An autonomous quadcopter follows the ground vehicle (left). The quadcopter's camera provides the operator a third-person view of the vehicle.

1 Introduction

The two key elements in user interfaces for teleoperating vehicles are 1) supporting efficient navigation through the remote environment and 2) providing the operator with a sense of presence or awareness of that environment. Traditionally, a video feed is used to provide information to the operator. This video feed, usually from an on-board, front-mounted camera, provides visually rich information of the environment. However, reported problems [6, 2, 10, 16, 21] are difficulties in navigation through the remote environment and a limited field of view [6], sometimes called "soda straw vision" [16].

To improve operator performance, many authors propose combining the video feed with range information to construct a 3D view [15, 4, 5]. Range information is obtained from on-board sensors that detect obstacles at distances. Map building software is then used to integrate these sensor data to provide a map that indicates the robots location in context of its surroundings. In contrast to video, such maps provide a wide-angle or 2D/3D view of the surroundings. However, in urban search and rescue (USAR) scenarios in disaster areas, using these maps has drawbacks. The maps 1) only show the robot's points of view; 2) are not real-time, and therefore do not support dynamic environments; and 3) require complex technology.

In this paper we address these drawbacks with an alternative approach in which we provide a live video feed with a third-person perspective (Fig. 1). An operator navigates a remote ground vehicle. Behind the ground vehicle flies a quadcopter with a downward facing camera providing the teleoperator with a live video feed of the ground vehicle and it's surroundings. We so achieve a telepresence third-person view, completely real [1], as opposed to the status quo that constructs an augmented reality view of the remote environment based on sensor data.

We conducted an experiment to evaluate our design by comparing operator performance to traditional and similar approaches. We adopted an search and rescue scenario with the aim of finding victims in a building. In the experiment, we asked participants to find victims using three interfaces for teleoperating the ground vehicle: the traditional front-mounted camera, a pole-mounted camera, and the flying camera. The results showed that the third-person views increased navigation efficiency and awareness of the environment.

The contributions of our work are as follows:

- An automonous flying camera for teleoperating a ground vehicle that provides a true real-time third-person, bird's eye perspective view.
- An experiment that demonstrates the advantages of a pole-camera design and a flying-camera design over a front-camera design in a simplified, simulated search and rescue scenario.

2 Related Work

A large body of work exists on user interfaces for teleoperating robots. In this paper we focus on operator experience navigating a single vehicle and do not discuss the complexities of user interfaces that involve team collaboration [14] or flying robots that aim to coordinate swarms of autonomous ground vehicles. Other related work addresses robot control paradigms, discussing the advantages

*e-mail:{daniel, vipul, sakamoto}@designinterface.jp
†e-mail:takeo@acm.org
‡e-mail:inami@inami.info

of deferred control through pointing destinations [19] or sketching paths [11]. In the described experiment, we test a real-time vehicle control paradigm using a joystick.

2.1 Thethered and Plan Views

Wickens and Hollands [24] describe a number of principle ways of presenting viewpoints that are useful to discuss the operator view on the remote robot. An immersed viewpoint shows the front view, and is captured by an egocentric camera mounted on the vehicle. Plan views are top-down views. Shiroma et al. [20] report of increased operator performance when providing a plan view captured by a wide-angle camera mounted on a pole on the remote vehicle. However, they also report of inflexibility of the pole and poor resolution due to the wide angle. Tethered views look at the vehicle from a third-person perspective. Furthermore, both plan and tethered views distinguish "north up" from "track up". In this paper, we are interested in "track up" tethered views in scenarios where there are no fixed cameras in the environment.

Hing et al. [10] describe a teleoperated airplane in which they construct a third-person view from a 3D model of the environment obtained in advance. The operator is presented with a "from-behind" view of the aircraft rendered in the virtual environment. The egocentric front camera is rendered perspectively correct in front of the plane. They make sure the virtual camera is y-up, independent of the roll of the airplane. Operators navigated the plane more fluently through the cluttered environment using this constructed external view than they did using the front camera. However, this approach is not realistic for our scenarios because it requires a 3D model of the environment and the precise location of the airplane in that model. Sugimoto et al. [21] describe a system that also constructs a from-behind view from an egocentric camera that does not need a pre-obtained virtual model of the environment. They use past video frames in which they draw a virtual representation of the vehicle at its calculated current position. Using the external view of the vehicle, operators performed better in precision navigation. However, neither solution described above can construct a bird's eye perspective.

Most described approaches that construct a third-person view, use some form of simultaneous location and mapping (SLAM). A 3D representation of the environment is constructed on the fly [15, 4, 5]. For instance Ferland et al. [5] use a depth sensing camera to construct a bird's eye perspective. However, this method is also limited to the environment that has been in line of sight with the ground vehicle, and it is not real time.

Nielsen and Goodrich [15] compare various (combined) representations of an egocentric video feed with low-fidelity sensor-constructed map data. They found that presenting the video and map in a combined 3D perspective view, made operators not only faster but also made them make fewer errors compared to video and or 2D representations. In a qualitative study, Drury [4] found that in a search and rescue scenario a video centric interface is preferred over a map-centric one for search activities.

None of the described prior art in teleoperating handles dynamic changing environments with a third-person, birds-eye view. Operators are provided with viewpoint constructed from an egocentric camera. In contrast, our proposed approach makes use of a live video feed.

2.2 Autonomous Flying Robots that Track a Target

Our teleoperator interaction builds upon a large body of work that involves tracking moving targets with cameras from above. Here, we review a few autonomous flying robots that track targets. Teulière et al. [23] describe a system in which a flying robot follows a car. They use color tracking, and anticipate occasions where the vehicle will be lost, such as when it travels under something. The project Flying Eyes by Higuchi et al. [9], and also Graether et al.

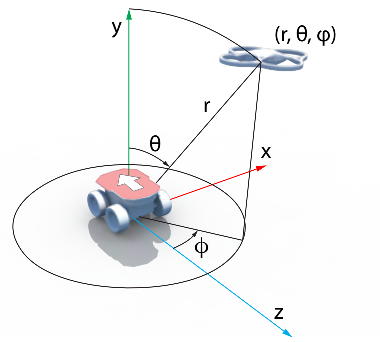

Figure 2: The proposed camera model. The coordinate system we use is spherical and is relative to the position and orientation of the ground robot for a track-up camera design.

[7] describe systems that track a human with an autonomous flying camera. They also use color tracking and provide a number of ways of tracking the subject. Although these projects use a flying camera to chase an object, the control of the object is not in the hands of the user, e.g. no teleoperation scenarios are explored. In addition, as far as we know, no motion stabilization is performed.

2.3 Cinematic Cameras

In computer games, third-person views are commonplace. A so-called "cinematic" camera [3] follows the game character through the 3D world guided by a constraint system borrowed from film-making theory, such as described by Katz [13]. Typical basic constraints include "align the camera up-vector with the world-up vector"; "maintain gaze at character"; "maintain distance and height relative to the character"; "maintain angle to the line-of-action". These constraints are solved using the 3D geometry of the virtual world. With a physical camera, we deal with additional constraints e.g. the camera cannot fly through walls and can also not teleport to provide a cut transition.

In addition, these cinematic cameras are animated according to content dependent functions such as zoom-in, zoom-out, change angle on character. For instance, certain car navigation systems relate their zoom on a vehicle to its speed and 3D computer aided design (CAD) software displays smooth transitions when switching views. These camera techniques are the inspiration for the proposed designs of the autonomous flying camera that controls a ground vehicle.

3 FLYING CAMERA DESIGN

The simplest design to provide an external view is to mount a camera on a pole attached to the rear of a ground vehicle. Then there is a 1:1 rigid connection between the ground vehicle and view, with no degrees of freedom, which implies that the vehicle is always in the center of the view. As shown in Fig. 2, we describe this camera using spherical coordinates. The distance r (the length of the pole) controls the apparent size of the ground vehicle in the captured image. The polar angle θ controls the viewing angle, and can change the operators view from a plan view (top down) to a from-behind view. The azimuth angle ϕ is fixed, relative to the ground vehicle, to provide a from-behind view. Additional mechanical control, such as using a linear actuator [22] to change the distance r, can add degrees of freedom. However, the distance has physical limitations with respect to the center of gravity of the vehicle and the turn radius. Further, low ceilings require special attention to avoid collisions.

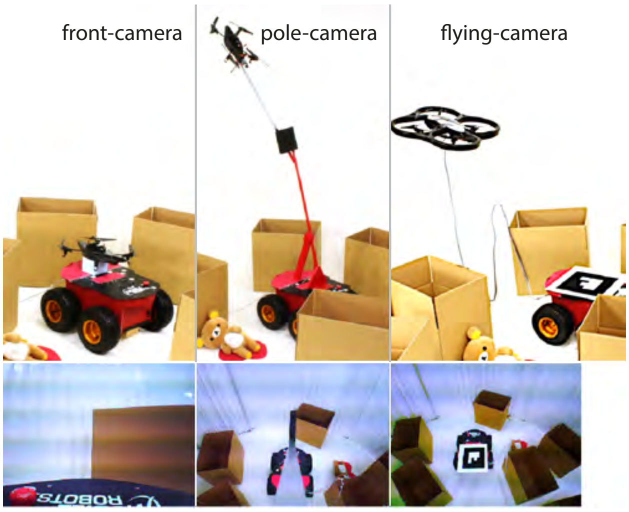

Figure 3: The three designs in the experiment. Shown on the left, a traditional front-mounted camera design in which an AR.Drone is mounted on top of the ground vehicle. The pole-mounted camera design is depicted in the center and has an AR.Drone mounted on an attached pole. Right, the flying-camera design. Here an AR.Drone is flying behind the ground vehicle. On top of the ground vehicle a fiducial marker is placed to estimate the position and orientation of the AR.Drone. The AR.Drone's camera is used in all conditions to achieve identical video quality and intrinsic camera settings. The bottom row pictures show the respective views as seen through the AR.Drone's camera.

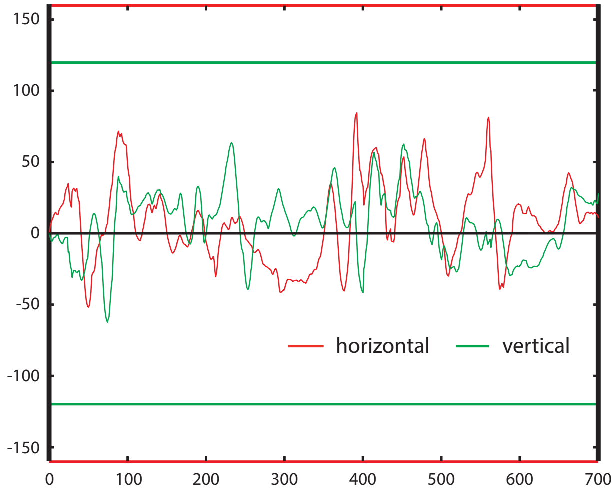

Figure 4: The performance of the ground vehicle tracking implementation. The ground vehicle is moving in a straight line with 0.19 m/s and its position is plotted as seen through the view of the flying camera. The horizontal axis depicts frames (15 fps). The vertical axis depicts the position of the center of the ground vehicle relative to the center of the camera in pixels (QVGA, 320×240 pixels). The horizontal difference is depicted in red, the vertical in green.

Therefore, our idea is to use a flying robot, specifically a quadcopter. All degrees of freedom in our camera system can be accommodated by altering the quadcopters's position and orientation relative to the ground vehicle. In addition to the looking at the ground vehicle as described in our camera model, a flying camera could add additional functionality for navigation assistance. For instance, it allows the teleoperator to "look around" or "scout" the environment. It can also provide a north-up view rather than a track-up one by using a compass. Because hovering flying robots are power hungry, we based our design on a thetered quadcopter to provide power through a power cable.

Unlike the pole camera design, the ground vehicle is not by definition in the center of the camera view. If the vehicle is lost, there are two possibilities. First case is that the vehicle moved out of the frame too quickly for the quadcopter to anticipate, in which case the quadcopter would maintain speed on the horizontal plane and increase altitude to zoom out. Second case is that the vehicle is lost within the frame because of occlusion. Then the quadcopter maintains the last known position and increases altitude, until the vehicle is found.

In addition, if the vehicle is not in the center of the camera view, we apply image-based motion stabilization to improve operator experience. The center of the vehicle is calculated in image space, and subtracted from the center of the screen. The image from the camera is then translated accordingly. In this way, the vehicle is always in the center of the display, and the inherent instability of a flying platform is compensated.

4 PROTOTYPE

Our prototype consists of standard available hardware. The ground vehicle is a basic model MobileRobots PIONEER 3 AT. The size of the vehicle is 0.52 by 0.48 by 0.28 meter ($l \times w \times h$). It is controlled through the vendor-supplied serial protocol over an IEEE 802.15.4 (ZigBee) wireless link.

The flying robot is a classic model AR.Drone [18], with a QVGA (320×240 pixels) resolution front camera that has a diagonal field of view of 90 degrees. We physically modified the tilt angle of the front facing camera to provide a bird's eye perspective. The camera looks downward at an angle of 60 degrees. The video of the camera is transmitted wirelessly (IEEE 802.11) through the vendor protocol and received with a frequency of 15 fps. The quadcopter is tethered with a power cable and is so powered through the ground vehicle. This gives us airtimes longer than the 10 minutes available with battery power.

We placed a 0.25 × 0.25 meter fiducial marker [12] on the ground robot to estimate the position and orientation of the quadcopter. The size proved to be adequate for the camera, distance, and lighting conditions for the experiment. As shown in Fig. 3, we used the marker for the earlier described camera model calculations. The video feed is received on a laptop, the ground station, on which the marker detection is performed. On the ground station we also perform image-based motion stabilization. The QVGA image received from the flying camera is inflated to VGA. The image is then positioned on a UXGA screen (1600×1200 pixels) so that the marker (ground vehicle) is in the center of the screen, using the 2D position of the marker in the received frame.

The quadcopter control consists of three independent control loops (proportional-derivative (PD)) to maintain distance and orientation to the ground vehicle. A transformation matrix is calculated using the fiducial marker and the camera intrinsics, in order to estimate the relative position and orientation of the quadcopter to the ground robot. The resulting navigation commands are sent back over the wireless link.

We measured the performance of the quadcopter camera system using the position of the center of the ground vehicle as seen in a video frame. As shown in Fig. 4, with the ground vehicle in motion on a straight path, the center of the ground vehicle stays well within the frame of the flying camera.

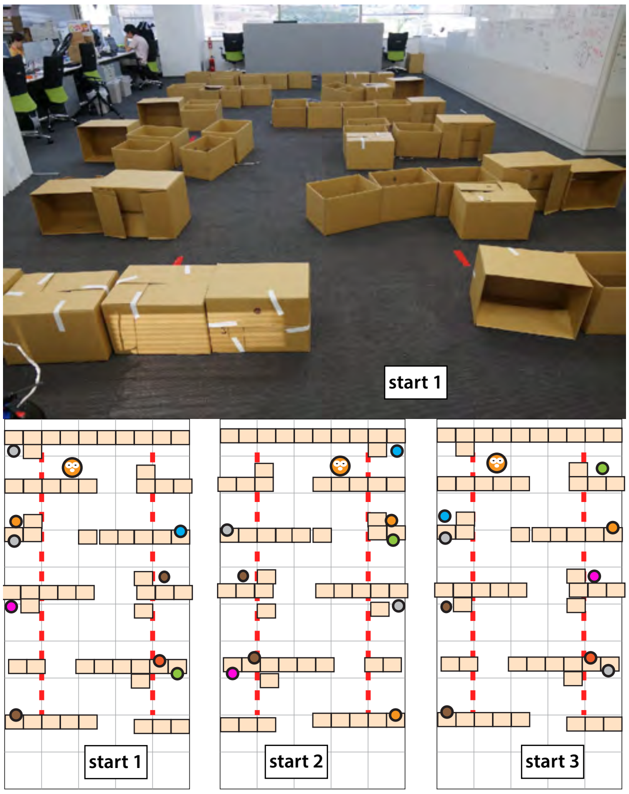

Figure 5: The search environment consist of a 5×10 meter space filled with cardboard boxes. The red dotted line may not be crossed with the ground vehicle, in order to give the quadcopter enough room for turning. In this space, we laid out three configurations of equal difficulty. In the photo, configuration one is shown after a mission; some boxes moved due to collisions.

5 EXPERIMENT

The goal of the experiment is to confirm the effectiveness of the proposed method both in efficient navigation and awareness of the environment. In a simulated search and rescue mission, we compared three camera designs: a traditional front camera, a simple pole camera and a flying camera. Our first hypothesis is that an external, bird's eye perspective (pole and flying cameras), performs better than a front-mounted camera. We expected that the front camera would provide less insight into the environment and would result in less efficient navigation. We assess good performance through 1) mission completion time, 2) number of victims found, and 3) number of collisions. Our second hypothesis is that operator performance achieved with the flying camera is comparable to that of the pole camera. We expected that the motion stabilization effectively counters the negative influence caused by the inevitable unstable motion of a lightweight quadcopter.

5.1 Experimental Design

We chose a within-subject design to reduce the error variance that we expected to find in a vehicle navigation performance task. Each participant completed three conditions in which the camera was the independent variable: traditional front camera, pole-mounted camera and flying camera. Since our task involved navigating through an environment and finding clues in that environment, we took care of carry-over effects by having participants visit the test environment prior to performing the missions, so that they could obtain sufficient environmental knowledge. Between conditions we changed the configuration of the environment as shown in Fig. 5, We also took counter measurements against vehicle navigation skill improvement: When vehicle speed was reduced to 0.19 m/s and rotation to 0.34 rad/s, navigating became easy. In addition, prior to their first mission we had participants perform a trial in which they practiced vehicle control.

We measured mission completion time, counted the number of victims found and counted the number of collisions with the environment. We also offered participants questionnaires before and after each mission. We asked for their experience in games and vehicle control first. As shown in Fig. 8, after each condition we asked them their confidence in finding all victims, their feeling of motion sickness, and to enter remarks. After they had completed the three conditions, we asked them to rank the conditions on appropriateness for the task accompanied with a written explanation. This subjective data is used in the discussion to explain the results.

5.2 Participants

Nineteen participants, 17 males and two females, were recruited from local universities to participate as teleoperators. Participants were students in computer science or human computer interaction, and ages ranged from 19 to 32. All participants had prior experience driving vehicles and playing computer games. The study duration was estimated at one hour, and participants were paid accordingly.

5.3 Procedure

Each participant started with a questionnaire regarding age, gender, prior experience in games and driving experience. The questionnaire was followed by a short practice session with the objective of having the participant gain familiarity with the vehicle dynamics and control in relation to the test environment. During the five-minute trial, the participant could see both the environment and the vehicle.

After the practice session, the participant was relocated to an adjacent room and instructed on the procedure. The three conditions were ordered semi-random and counter balanced. All permutations of the conditions were distributed over the participants. With 19 participants, each sequence of conditions occurred three times except the first one, which occurred four times. Each condition was followed by a few questions regarding the mission. They rated their performance and motion sickness on a 7-point Likert scale. In addition the questionnaire contained open space for comments.

Prior to each mission, the search environment was prepared according to the layout shown in Fig. 5. All three conditions were piloted and found to be completed within five minutes. A concluding questionnaire contained additional questions regarding the three conditions, asking which condition they preferred and why.

5.4 Task

A known but cluttered environment has to be searched to find victims. The participant does not know the number of victims. The task and environment are similar to Nielsen and Goodrich [15] and Drury et al. [4] and consists of navigating the remote vehicle through a path: "go through the maze and find all the victims. Be thorough, but do not waste time and avoid collisions." We asked participants to report found victims to an observing researcher by pointing to the screen.

5.5 Prototypes

The front camera of the quadcopter was used in all three conditions. It ensured identical video quality as well as identical camera intrinsics between conditions. In the front camera condition, the quadcopter is mounted on the ground vehicle. We made sure that the camera shows the front corners of the robot, which we deemed essential for distance-to-vehicle estimation. In the pole condition,

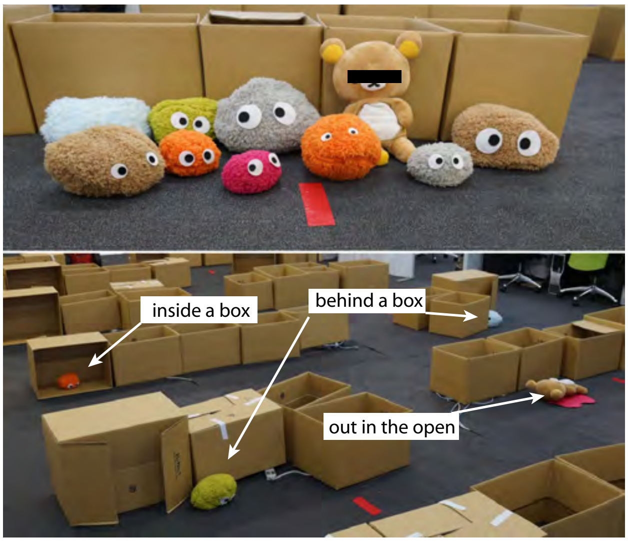

Figure 6: The ten victims in the mission. They vary in size and color, but from the eyes they are easily recognized as such. As shown, victims were placed inside a box, behind a box, or out in the open according to Fig. 5

we mounted the quadcopter on a pole behind the robot, equally distanced and angled as in the flying robot condition.

The AR.Drone has obstacle avoidance safety measurements built into its firmware. These measurements interfered negatively with our camera control loop. From time to time, the quadcopter got confused by the height difference when it flew over obstacles. In these cases a skilled operator took over control and navigated the quadcopter back into position. These interventions were hidden from the participant. We chose this wizard of Oz method to prevent breaks that would occur when stopping and restarting the experiment.

The quadcopter was powered through a separate battery pack mounted on the ground vehicle. Two additional battery packs and chargers for both the flying and ground robots assured enough power in consecutive missions.

5.6 Teleoperating Setup

In a separate room, the teleoperator was positioned at a desk with a 19-inch monitor. The eye-to-screen distance was approximately 0.75 meter. As shown in Fig. 1, the teleoperator wore active noise-canceling headphones (Audio Technica Quiet Point) playing music so that the physical robots could neither be seen nor heard. The soundtrack of the movie Inception (Nolan, 2010) was played to elevate the operator's sense of urgency.

The ground vehicle was controlled using a Logicool gamepad with two analog joysticks. Rotation was assigned to the left joystick (left, right); speed (forward, backward) was assigned to the other. The screen showed the camera view with the ground robot centered in the view as described in the prototype section.

5.7 Environmental Setup

Our environment consists of a 5×10 meter space filled with cardboard boxes. The boxes are higher than the ground robot, but lower than the flying and pole-mounted cameras. As shown in Fig. 5, three paths of equal level of difficulty level were outlined. In this environment ten colorful soft toy animals of various sizes were distributed in three different ways as shown in Fig. 6. All toy animals (victims) were placed with eyes facing outwards. Victims inside a box were generally easier to see from a low angle, whereas victims located behind a box were generally easier to spot from a higher angle.

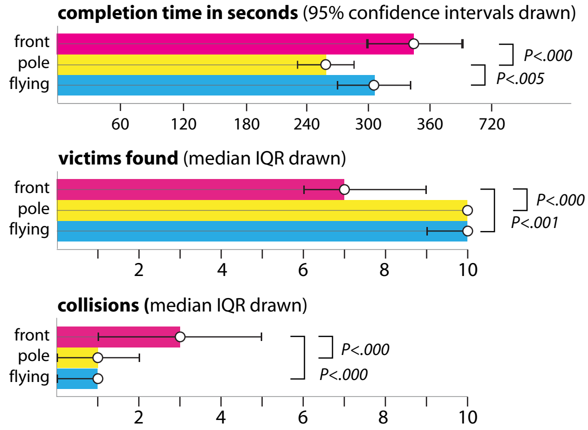

Figure 7: The results of the experiment. The completion time for the pole condition is significantly faster than the front and flying condition. The front condition performs significantly worse on victims found as well as on collisions with the environment.

6 RESULTS

All participants completed the missions within nine minutes in all conditions. On a few missions, 5 out of 57, the task had to be briefly stopped due to technical errors, but was restarted within three minutes. These missions were included in the end results but we excluded the duration of the break from the mission completion time.

First, we analyzed the data for carry-over effects, e.g. whether the order of the missions yielded significant differences. We ordered the data to compare the first, second and third missions. A repeated measures ANOVA determined no significant differences in completion time ($F(2,54) = .313, P = .733$). A Friedman Test revealed no statistically significant differences between the number of collisions ($\chi^2(2) = .393, P = .821$) nor between the number of victims found ($\chi^2(2) = 1.536, P = .464$). That means our assumption was correct: no carry-over effects in the mission order were found.

6.1 Completion Time

A repeated measures ANOVA determined that completion time differed significantly between the conditions ($F(2,54) = 14.33$, $P < .0005$). Post-hoc tests using the Bonferroni corrections revealed a significant difference between the pole and the front (pole<front, P<.000), flying condition (pole<flying, P<.005). However, no significant difference was found between the front and flying conditions (front<flying n.s.) as shown in Fig. 7

6.2 Number of Victims Found

A Friedman Test showed that the number of victims found differed significantly between the conditions ($\chi^2(2) = 25.964, P = .000$). A post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $P < .017$. There were no significant differences found between the pole and flying conditions ($Z = -.905, P = .366$). However, there was a statistically significant reduction in the number of victims found in the pole vs. front ($Z = -3.541, P = .000$) and flying vs. front ($Z = -3.429, P = .001$).

This result is accordance with our expectations that a third-person view through the bird's eye perspective would allow a better view of the environment.

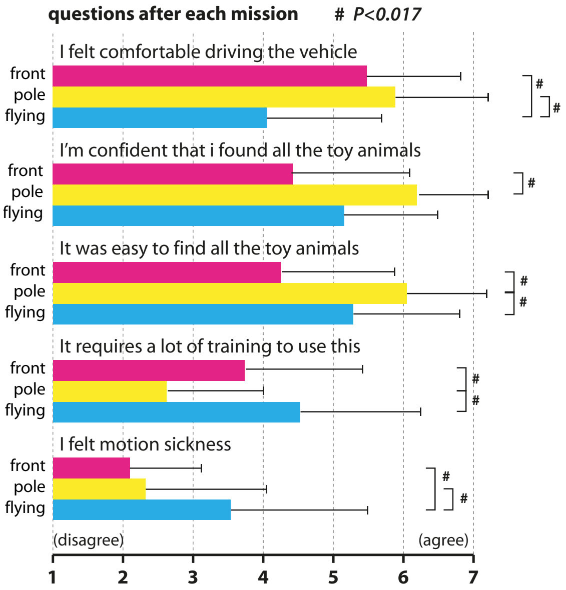

Figure 8: Following each mission, we asked participants to answer the questions shown in the figure on a 7-point Likert scale. Means and Std. Deviations are indicated per conditions, and a # depicts a significant difference.

6.3 Collisions

A Friedman Test showed that the number of collisions differed significantly between the conditions. ($\chi^2(2) = 26.984, P = .000$) Similar to the number of victims found, a post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Bonferroni correction applied, resulting in a significance level set at P<.017. There were no significant differences between the pole and flying conditions ($Z = -.707, P = .480$). However, there was a statistically significant reduction in collisions found in the pole vs. front ($Z = -3.619, P = .000$) and flying vs front ($Z = -3.643, P = .000$).

This result is as expected. Despite that the front of the ground vehicle was shown in the front camera view, operators seemed to have difficulty in judging the distance to the environment.

6.4 Qualitative Analysis

When rating the three conditions, 15 out of 19 participants selected the pole condition as the best for this task. Three selected the flying condition and only one selected the front condition. Out of the 19 participants, 15 picked the front camera as the least suitable condition: the other four selected the flying camera.

In their remarks, participants preferred the quadcopter's high elevation but disliked its slow turning. A few participants suggested additional controls for the flying camera to have it look sideways. Despite the implemented motion stabilization, a number of participants reported the poor stability of the flying robot, and some attributed this directly to their navigation performance. The observing researchers noticed that in the flying condition the participants acted more carefully.

The after-each-mission questions, shown in Fig. 8, were rated on a 7-point Likert scale. Post-hoc analysis with Wilcoxon Signed-Rank Tests was conducted with a Bonferroni correction applied, resulting in a significance level set at P<0.05. As shown, the pole-condition received the best scores, participants found this condition significant easier to find the toy animals and also thought it required the least amount training to master. Participants reported a significant stronger feeling of motion sickness and a significant reduced feeling of comfortable driving in the flying condition.

Although all conditions used identical cameras, a number of participants reported that they experienced the field of view in the front-mounted condition as too narrow. Some participants argued for additional side cameras, and a rear-camera view when going backwards.

6.5 Discussion on the Results

The experimental results show that third-person, bird's eye views provided by the pole and flying camera condition performed better on number of victims found and number of collisions with the environment. This result is in line with improved operator performance reported in prior art. However, the results also show that the flying camera condition was not as efficient as the pole camera condition in mission completion time. On the basis of observation during the missions and the feedback from the questionnaires, this seems to be caused by two reasons. First, it took time for the flying camera to reach the appropriated position after the user turns the direction of the ground vehicle, which makes, for instance, "peeking" into a box slow. Second, the instability of the flying camera caused more careful navigations. We hope that a robust quadcopter implementation will improve completion time to the level of the pole camera in the future.

7 LIMITATIONS AND FUTURE WORK

A limitation of our experiment is that we could only test the basic functionality of our system, because the task was intense and the participants were inexperienced in teleoperating vehicles with flying cameras. Further studies, when operators gain experience, are better suited to test these extended camera functions.

7.1 Additional Controls

With a flying camera we can introduce additional controls, for which a few participants expressed desire. For instance, a flying camera makes it possible to switch from a track-up to a map-up view, which can be useful for some control scenarios. A zoom-in/zoom-out function could be implemented by adjusting the distance to the ground vehicle, similar to how it is implemented in CAD software. Also, we have not explored flying camera interfaces with point-and-click or sketch-based control paradigms.

We have started to implement a "look-around" function. In principle we distinguish a "vehicle-centered" look around function from a "camera-centered" look around one. Whereas the former could be achieved by rotating the ground vehicle while keeping the ground vehicle in the center of the view, the latter rotates the quadcopter and so moves the ground vehicle out of the center of the view. We implemented a quadcopter centered rotation using a single button control that behaves similar to peeking" at the side mirror of a car: the view rotates and returns back to the default view on the ground vehicle. In this way, additional camera commands can be implemented without adding the burden of navigating an additional vehicle.

7.2 Manipulation Task

In this paper we focused on a search task, so one might claim that a flying robot without a ground vehicle might suffice, a limitation of the proposed system. However, the goal of this work is to support a search and rescue scenario in which the ground robot has a manipulation task in addition to the search task. In the future we hope to extend our research with manipulation tasks. For instance, such as described by Hashimoto et al. [8] who currently uses a static camera or Zhu et al. [25] where the user currently manually controls both the camera and a robot arm.

Figure 9: The preliminary implementation of a balloon-mounted camera. Three winches control the distance to the ground vehicle and provide the teleoperator with zoom functionality as shown on the left.

7.3 Balloon

A number of participants indicated the need for a higher camera to improve operator performance through a better overview. For outdoor scenarios, a balloon-mounted camera might provide a good solution. A camera mounted on an aerostat balloon, or airship, as shown by Okura et al. [17], provides an effective way to provide a plan view. A balloon doesn't have the weight of a pole, so it could potentially go very high using a winch for zoom. A moving ground robot will drag the balloon and will cause a change in azimuth from plan to from behind. The natural inertia provides a smooth transition. In order to maintain gaze at the ground vehicle, the balloon is mounted through two wires. Fig. 9 shows our prototype implementation.

8 Conclusion

We presented the first steps towards an autonomous flying-camera system for improving the teleoperator's interface to remote ground vehicles. A flying-camera tracks the ground vehicle that the teleoperator controls and provides the operator with a motion-stabilized external, third-person perspective view on the ground vehicle.

We conducted an experiment to compare a flying-camera system with a traditional front-mounted camera system and a pole-mounted camera system in a simulated, teleoperated USAR scenario. The results showed that the third-person perspective provided by our flying-camera and pole-mounted camera resulted in fewer collisions of the ground vehicle with the environment and more victims being located compared to the front-mounted camera. The pole-camera had the fastest mission completion time, but results indicate that an improved implementation of the flying camera will result in equal performance. In addition, a flying camera has more flexibility than the pole-mounted camera.

Acknowledgements

The authors wish to thank Yoshimasa Furukawa for translating the Japanese questionnaires.

References

[1] R. Azuma. A survey of augmented reality. *Presence*, 6(4):355–385, Aug. 1997.

[2] W. Boonsuk, S. Gilbert, and J. Kelly. The impact of three interfaces for 360-degree video on spatial cognition. In *Proc. CHI '12*, pages 2579–2588, 2012.

[3] S. M. Drucker and D. Zeltzer. Camdroid: a system for implementing intelligent camera control. In *Proceedings of the 1995 symposium on Interactive 3D graphics*, I3D '95, pages 139–144, 1995.

[4] J. L. Drury, B. Keyes, and H. A. Yanco. Lassoing hri: analyzing situation awareness in map-centric and video-centric interfaces. In *Proc. HRI '07*, pages 279–286, 2007.

[5] F. Ferland, F. Pomerleau, C. T. Le Dinh, and F. Michaud. Egocentric and exocentric teleoperation interface using real-time, 3d video projection. In *Proc. HRI '09*, pages 37–44, 2009.

[6] M. Fiala. Pano-presence for teleoperation. In *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 3798–3802. IEEE, 2005.

[7] E. Graether and F. Mueller. Joggobot: a flying robot as jogging companion. In *Proc. CHI EA '12*, pages 1063–1066, 2012.

[8] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi. Touchme: An augmented reality based remote robot manipulation. In *Proc. ICAT '11*, 2011.

[9] K. Higuchi, Y. Ishiguro, and J. Rekimoto. Flying eyes: free-space content creation using autonomous aerial vehicles. In *Proc. CHI EA '11*, pages 561–570, 2011.

[10] J. T. Hing, K. W. Sevcik, and P. Y. Oh. Development and evaluation of a chase view for uav operations in cluttered environments. *J. Intell. Robotics Syst.*, 57(1-4):485–503, Jan. 2010.

[11] T. Igarashi, R. Kadobayashi, K. Mase, and H. Tanaka. Path drawing for 3d walkthrough. In *Proc. UIST '98*, pages 173–174, 1998.

[12] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999. (IWAR '99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94, Washington, DC, USA, 1999. IEEE Computer Society.

[13] S. D. Katz. *Film Directing Shot by Shot: Visualizing from Concept to Screen*. Michael Wiese, 1991.

[14] B. Larochelle, G. Kruijff, N. Smets, T. Mioch, and P. Groenewegen. Establishing human situation awareness using a multi-modal operator control unit in an urban search amp; rescue human-robot team. In *RO-MAN, 2011 IEEE*, pages 229–234, 2011.

[15] C. W. Nielsen and M. A. Goodrich. Comparing the usefulness of video and map information in navigation tasks. In *Proc.HRI '06*, pages 95–101, 2006.

[16] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks. Ecological interfaces for improving mobile robot teleoperation. *Trans. Rob.*, 23(5):927–941, Oct. 2007.

[17] F. Okura, M. Kanbara, and N. Yokoya. Augmented telepresence using autopilot airship and omni-directional camera. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 259–260, 2010.

[18] Parrot USA. Ardrone. `http://ardrone.parrot.com/`.

[19] T. Sekimoto, T. Tsubouchi, and S. Yuta. A simple driving device for a vehicle implementation and evaluation. In *Intelligent Robots and Systems, 1997. IROS '97., Proceedings of the 1997 IEEE/RSJ International Conference on*, volume 1, pages 147 –154 vol.1, sep 1997.

[20] N. Shiroma, N. Sato, Y. huan Chiu, and F. Matsuno. Study on effective camera images for mobile robot teleoperation. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pages 107–112, 2004.

[21] M. Sugimoto, G. Kagotani, H. Nii, N. Shiroma, M. Inami, and F. Matsuno. Time follower's vision: A teleoperation interface with past images. *IEEE Comput. Graph. Appl.*, 25(1):54–63, Jan. 2005.

[22] S. Takei, M. Iida, and T. Naemura. Kinereels: extension actuators for dynamic 3d shape. In *Proc. SIGGRAPH '11 Posters*, pages 84:1–84:1, 2011.

[23] C. Teuliere, L. Eck, and E. Marchand. Chasing a moving target from a flying uav. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4929 –4934, sept. 2011.

[24] C. Wickens and J. Hollands. *Engineering Psychology and Human Performance*. Prentice Hall, 1999.

[25] D. Zhu, T. Gedeon, and K. Taylor. Exploring camera viewpoint control models for a multi-tasking setting in teleoperation. In *Proc. CHI '11*, pages 53–62, 2011.