2017 IEEE International Symposium on
Safety, Security and Rescue Robotics (SSRR)
Shanghai, China, October 11-13, 2017

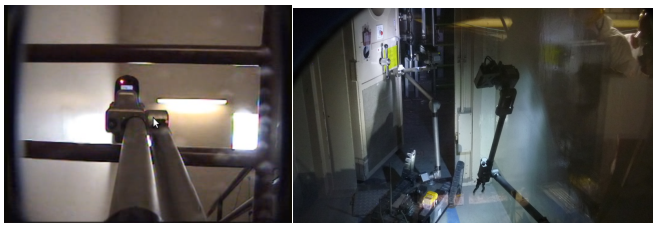
Visual Servoing for Teleoperation Using a Tethered UAV

Xuesu Xiao[1], Jan Dufek[1] and Robin Murphy[1]

Abstract— **This paper presents a visual servoing approach for robotic teleoperation using a tethered unmanned aerial vehicle (UAV). When teleoperating a robot, human operator's perception of the remote situation is limited by the robot's onboard camera. This deteriorates situational awareness and poses challenges on operation precision and efficiency. Teleoperated visual assistants are used in practice. For example, in Fukushima Daiichi nuclear disaster decommissioning, a secondary ground robot is used to follow and watch the primary robot. However, this requires two robots and 2-4 operators to perform one task. Furthermore, it introduces more problems, such as extra teamwork demand, miscommunication risk, suboptimal viewpoints. This work proposes to use a tethered UAV to replace the extra ground robot and human operators. In order to visually assist the primary robot autonomously, a visual servoing algorithm is developed and implemented based on a fiducial marker mounted on the primary robot, representing the operator's point of interest. Visual servoing configuration is controlled using 6 Degrees of Freedom of the fiducial. Servoing performances from physical experiments are analyzed. This paper lays the groundwork for and points out the direction of further visual assisting research.**

I. Introduction

Robots are widely used in DDD (Dangerous, Dirty, and Dull) environments, where human presence is extremely difficult or impossible [1]. Full autonomy is desirable for those robotic missions. However, for mission-critical tasks, human are still in the loop due to technological limitations from sensing, computing, intelligence, and safety. Therefore projecting human presence to remote environment [2] is still a powerful application of mobile robots, which leverages current technologies and actual field demand. One of the challenges associated with robotic teleoperation is the perception limitation of human operators caused by robot's onboard sensors. The onboard camera, for example, can only provide the operator with a first person view (FPV). FPV doesn't include any depth information and the perception is constrained by the camera's field of view. As shown in Fig. 1a, the operator is not able to tell if the radiation sensor held by the manipulator arm has reached the grill. To address this problem, multiple cameras are installed onto the robot, to provide multiple visual feedback of different functional units. However, switching constantly between different view points disrupts the operator and blind spots still remain in the environment which are out of multiple stationary cameras' field of view. A common practice for ground robots in homeland security applications, disaster response, and inspection tasks is to use a secondary robot providing a view of task being performed by a primary robot. In Fukushima Daiichi nuclear power plant, teleoperated robots are used in pairs from the beginning of the response to reduce the time it takes to accomplish a task. Fig. 1b shows two iRobot Packbots being used to conduct radiation surveys and read dials inside the plant facility, where the second Packbot provides camera views of the first robot in order to manipulate door handles, valves, and sensors faster. Other examples include QinetiQ Talon Unmanned Ground Vehicles (UGVs) letting operators see if their teleoperated Bobcats end loader bucket had scraped up a full load of dirt to deposit over radioactive materials. Problem with this approach, however, is that it takes two robot and 2-4 operators to perform one single task. The two sets of robot operators find it difficult to coordinate with the other robot in order to get and maintain the desired view but a single operator becomes frustrated trying to operate both robots. The extra teamwork demand and miscommunication may lead to problems: in 2014, an iRobot Warrior costing over $500K was damaged due to inability to see that it was about to perform an action it could not successfully complete. In addition to the huge economic loss, the 150 kg robot was too heavy to be removed without being dismantled and thus cost other robots times and increase their risk as they had to navigate around the carcass until another robot could be modified to dismantle it.



(a) FPV of Onboard Camera (b) Two Packbots opening a Door

Fig. 1. Problems with Current Robotic Teleoperation

In this research, an Unmanned Aerial Vehicle (UAV) tethered to the primary robot is used to replace the second visual assisting robot and the extra operators. The visual assisting process is implemented using a tether based controller for positioning and a visual servoing algorithm to keep looking at the operator's Point of Interest (POI) at a constant 6-DOF pose.

This paper is organized as follows: first, related work is reviewed in Sec. II. Second, the heterogeneous robotic

[1]Xuesu Xiao, Jan Dufek, and Robin Murphy are with Department of Computer Science and Engineering, Texas A&M University, College Station, Texas 77843 {xiaoxuesu, dufek}@tamu.edu, murphy@cse.tamu.edu

team is introduced in Sec. III. In Section IV, the visual servoing and control algorithm are presented. Experiments results and analysis are shown in Sec. V. As this paper lays the ground work for further research on visual assisting for teleoperation, Sec. VI discusses about future work. Sec. VII concludes the paper.

II. Background

A. Visual Servoing

Visual servoing, also known as vision based robot control, is a technique to control the motion of a robot using feedback information extracted from a vision sensor, or visual feedback. The idea of visual servoing originates from the 80's [3]. The original idea of visual servoing is for robotic manipulation. It could be subdivided into two categories based on the configuration of the robot end effector and camera: eye-in-hand [4] and eye-to-hand [5]. In the former approach, camera is attached to the moving hand and observes the relative position of the target. This corresponds to the onboard camera of the ground robot but assumes the motion of the manipulator arm is automated. The second approach is to fix the camera in the world and observe the target and the motion of the hand. The traditional eye-hand visual servoing has been extended beyond the manipulation field. [6] used a ceiling mounted camera to control mobile robots. Using the camera's visual feedback, [7] controlled a hyper redundant snake robot as controlling a differential drive car. A UAV was able to chase a moving target based on all in-plane visual clues in [8]. [9] and [10] used a UAV to visually navigate an unmanned aerial vehicle to to rescue drowning victims. Our proposed approach, however, is dynamically placing the camera in the world, which is the camera's 6 Degrees of Freedom (DOF) configuration space. In terms of control techniques, Image-based (IBVS), Position/pose-based (PBVS), and hybrid approach are the three main types [4] [11]. IBVS controls the robot only using the error between current and desired features in the image frame. Target pose is not estimated [12]. In PBVS, image features are used to calculate the pose of the object of interest, and then the robot's motion commands are generated based on the pose. This approach is servoing in 3D instead of in image frame. In our work, we augmented the classical PBVS by 3 additional dimensions: the robot is able to servo any configurations in the target's whole 6-DOF configuration space, three translational and three rotational components.

B. AprilTag

In the pursuit of retrieving the 6-DOF configuration of the servoing target, artificial features (fiducials) could largely increase the tracking precision and robustness and reduce the computational cost. More importantly, extra information could be computed to assist better visual servoing, such as scale, depth, and rotations. Although the environments where our heterogeneous robotic team works in does not include any engineered features, the visual servoing is happening around the primary ground robot. Features could be added on different functional components of the primary robot, for example, on the manipulator arm to push a door, on the gripper to pick up an object, on the driving components to assess passability, etc. AprilTag [13] is an appropriate artificial feature for this purpose. It provides 6-DOF configuration of the tag, using only one frame of one singe camera. The pose of the tag estimated by the AprilTag system contains scale information with respect to the scale of the tag. Since the tag size could be designed beforehand, the visual servoing controller issues motion commands which is not up to scale (no universal scale ambiguity). Other benefits are depth perception and the extra three rotational components in the configuration space.

Teleoperation mission involves a variety of tasks. A good servoing system is more than simply keeping the Point of Interest in the field of view. The extra rotational components, along with the accurate translation including depth, makes it possible that the object of interest is presented in an optimal pose in the visual feedback. For example, when operating a gripper to pick up an object, the operator may prefer perceiving both in-plane and depth information at the same time, to simply seeing the gripper from an arbitrary angle in space. In this paper, all length units are in AprilTag units (one AprilTag is 2 unit $\times$ 2 unit, 1 unit = 8.5 cm) and all angular units are in radians.

C. Tethered UAV

UAVs are widely used in disaster response and recovery phase [14]. The outstanding mobility and relatively low cost make UAVs more and more popular in the safety, search and rescue community. UAV with a tether seems to deteriorate the aircraft's superior mobility at first glance. However, depending on the particular application, tether may be beneficial for multiple reasons. For example, in [15] researchers developed a stabilization algorithm for a flying vehicle on a taut tether only using inertial sensing, which is justified by the low cost and more friendly user-vehicle interaction, especially for novice operator. [16] used a tether to achieve high-speed, steady flight in a confined environment. Another advantage provided by the tether is lower weight (no battery onboard) and increased flight duration (power from ground).

For our visual assisting purpose, the benefit of using a tethered UAV in indoor cluttered environments is three-fold: 1. it matches the battery duration of UGV and UAV. UGV's battery could last hours because of its large payload capacity. Conventional untethered small UAV, however, can only fly for about 15 minutes without recharging. Using a tether for power transmission, the UAV's flight time could be largely increased and thus matches with its ground partner. 2. Better positioning and stabilization could be achieved using the tether in a GPS-denied environment. Although indoor visual odometry is a potential solution, it is still not robust enough, and requires binocular vision and large computational power. 3. In the case of a malfunction or accident, the tether could be used to retrieve the UAV if the UGV is still mobile.

Fig. 2. Heterogeneous Robot Team

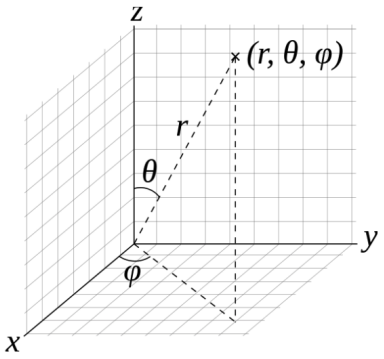


Fig. 3. Vehicular Position Coordinates

III. System Overview

Our heterogeneous robotic team is composed of a UGV, iRobot Packbot, and a tethered UAV, Fotokite Pro (Fig. 2 left). Fig. 2 right shows a proof-of-concept scenario where Fotokite is assisting Packbot to pick up a radiation sensor and drop it into the white vertical pipe.

A. Primary Robot

iRobot Packbot is used as primary agent for teleoperation missions. Packbot has been used in different scenarios, including Iraq and Afghanistan wars, searching the debris of World Trade Center after 9/11 in 2001, and Fukushima Daiichi nuclear plant after 2011 Tohoku earthquake and tsunami. It has at least three onboard cameras, which, however, are still not sufficient for good situational awareness (Fig. 1a). In confined and cluttered search and rescue mission, Packbot is always deployed in pairs (Fig. 1b) and requires 2-4 operators. Our system is aiming at automating the visual assisting process using a secondary tethered UAV. Packbot operator is provided a third person view of the situation for different ongoing tasks during the mission, for example, opening a door, turning a valve, picking up an object, or driving through a narrow passage, which have different optimal teleoperation view points.

B. Visual Assistant

Fotokite [17] is used as visual assistant. The UAV's tether angle sensor and ground station's tether reel encoder make localization and positioning with respect to its master ground robot possible, without GPS signal and visual odometry. The controls for Fotokite is based on tether angle, length and inertial measurement [15]. The vehicle position control uses *tether length* r, *elevation* θ (vertical angle of tether), and *azimuth* ϕ (horizontal angle of tether). The position of the vehicle could be represented in polar coordinate system (Fig. 3) and is easily transformed to Euclidean space:

$$\begin{cases} x = & rsin\theta cos\phi \\ y = & rsin\theta sin\phi \\ z = & rcos\theta \end{cases} \tag{1}$$

The rotational components of the vehicle (yaw, pitch, and roll) are represented with respect to the initial frame by a quaternion. Fotokite is equipped with a gimbal which has its own separate pitch and roll. For our visual assisting purposes, the camera's rotations need to be considered. So the rotational components of the visual assistant is the vehicular *yaw*, the camera's *pitch* and *roll*. The visual assistant's configuration space could be represented as $\begin{bmatrix} x, & y, & z, & yaw, & pitch, & roll \end{bmatrix}^T$. The onboard controller stabilizes this 6 DOF and waits for further motion command.

C. Setup

The ground station of Fotokite is mechanically mounted onto Packbot. At this stage of the project, automatic landing and take off has not been implemented yet. Since the visual assistant only provides feedback to the primary robot's operator, there is no need of communication between the two agents. A server is built as a relay between the ground station and off-board visual assisting control unit, which talks to each other via 2.4GHz radio. The server receives status updates from Fotokite ground station's serial port and passes them wirelessly to the control unit. Vehicle motion commands are computed at the control unit, transmitted to the server by radio, and sent to the ground station. Fotokite camera's video is streamed wirelessly by an onboard media encoder and streamer.

IV. Visual Servoing

In this section, the visual servoing algorithm is presented, which takes in the live video stream of the visual assistant's camera, and issues motion commands to servo the Point of Interest in the camera's 6-DOF configuration space.

A. Three Coordinates Systems

The three coordinates systems used in the visual servoing are illustrated in Fig. 4. The ground station is fixed to the primary robot and is defined as the inertial frame. For convenience, horizontal plane is defined as zx plane, while y axis is pointing up vertically. This easily aligns with the camera frame. *Tether length*, *elevation*, and *azimuth* are defined in the inertial frame. The position of the Fotokite $\begin{bmatrix} x_f, & y_f, & z_f, \end{bmatrix}^T$ is determined by Eqn. 1 after reorienting the axes. Since the translation from the vehicle Center of Mass (COM) to the camera origin is negligible, the

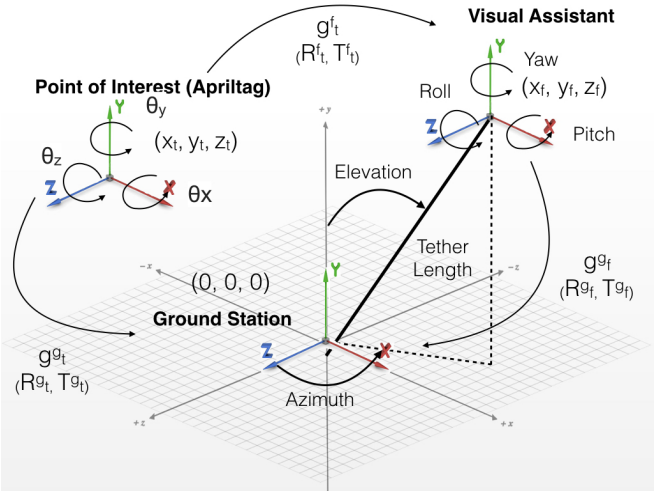


Fig. 4. Three Coordinate Systems

vehicle frame and camera frame is treated equivalently. The rotational components $\begin{bmatrix} yaw, & pitch, & roll \end{bmatrix}^T$ is the rotation angle with respect to the y, x, and z axis, respectively. By the same token, the POI (AprilTag) frame is defined with respect to the camera frame. The homogeneous transformations from Fotokite to ground station, and from tag to Fotokite, are respectively defined as:

$$\mathbf{g_f^g} = \begin{bmatrix} \mathbf{R_f^g} & \mathbf{T_f^g} \\ 0 & 1 \end{bmatrix}, \quad \mathbf{g_t^f} = \begin{bmatrix} \mathbf{R_t^f} & \mathbf{T_t^f} \\ 0 & 1 \end{bmatrix} \tag{2}$$

where $\mathbf{R}$ and $\mathbf{T}$ denote the rotation matrix and translation vector.

Assuming a point in the AprilTag frame to be $\bar{\mathbf{q}}_\mathbf{t} = \begin{bmatrix} x_t, & y_t, & z_t \end{bmatrix}^\mathbf{T}$, we could apply the following coordinate system transformation to express it in the ground station frame, where $\mathbf{g_f^g}$ could be derived by the current flight status and $\mathbf{g_t^f}$ is given by the AprilTag tracking system:

$$\bar{\mathbf{q}}_\mathbf{g} = \mathbf{g_t^g} \cdot \bar{\mathbf{q}}_\mathbf{t} = \mathbf{g_f^g} \cdot \mathbf{g_t^f} \cdot \bar{\mathbf{q}}_\mathbf{t} \tag{3}$$

In order to observe the POI from a fixed pose, independent of how the POI moves in the free space, our visual servoing controller should maintain a constant $\mathbf{g_t^f}$, the homogeneous transformation from the POI to Fotokite frame. We denote this desired transformation as $\mathbf{g_t^f}*$. The controllable states in the system are $\begin{bmatrix} x, & y, & z, & yaw, & pitch, & roll \end{bmatrix}^T$, which determines $\mathbf{g_f^g}*$, our desired vehicle configuration. So we have an alternative way to express $\bar{\mathbf{q}}_\mathbf{g}$:

$$\bar{\mathbf{q}}_\mathbf{g} = \mathbf{g_t^g} \cdot \bar{\mathbf{q}}_\mathbf{t} = \mathbf{g_f^g}* \cdot \mathbf{g_t^f}* \cdot \bar{\mathbf{q}}_\mathbf{t} \tag{4}$$

Combining Eqn. 3 and Eqn. 4 the desired vehicle configuration could be calculated:

$$\mathbf{g_f^g}* = \mathbf{g_f^g} \cdot \mathbf{g_t^f} \cdot {\mathbf{g_t^f}*}^{-\mathbf{1}} \tag{5}$$

$\mathbf{g_f^g}*$ could be further decomposed to $\mathbf{R_f^g}*$ and $\mathbf{T_f^g}*$, from which the controls to Fotokite could be derived.

B. Point of Interest Pose Estimation

$\mathbf{g_t^f}$ is computed by the AprilTag tracking system. 6-DOF configuration of the tag could give the homogeneous transformation. Based on different teleoperation tasks, different desired observing position and orientation could be easily defined. For simplicity, in this paper the desired observing pose is chosen as zero rotation and 10 unit distance away from the camera origin along the z axis. As a result, the AprilTag will always locate in the middle of the image frame, facing straight toward the camera.

C. Vehicular and Camera Motion Control

The status updates (*tether length* r, *elevation* θ, *azimuth* ϕ, and quaternion representing the vehicle orientation) of the current sensed vehicular configuration is used to compute $\mathbf{g_f^g}$. The gimbal pitch and roll, however, need to be estimated using integration since it is not provided by the current version of Fotokite SDK firmware.

Using $\mathbf{g_f^g}$, along with $\mathbf{g_t^f}$ from AprilTag and predefined $\mathbf{g_t^f}*$, $\mathbf{g_f^g}*$ is computed from Eqn. 5. After computing the desired transformation between the vehicle and ground station, $\mathbf{g_f^g}*$ is translated into the visual assistant's configuration space: $\begin{bmatrix} x*, & y*, & z*, & yaw*, & pitch*, & roll* \end{bmatrix}^T$. The first four dimensions are controllable by the vehicle, while the gimbal is responsible for the last two. $\begin{bmatrix} x*, & y*, & z* \end{bmatrix}^T$ need to be represented in the form of $\begin{bmatrix} tehter\ length*, & elevation*, & azimuth* \end{bmatrix}^T$. After this process, the interrelated state space dimensions are decoupled into 6 independent variables. 6 PID controllers are used to drive those 6 independent variables to the desired value.

Given the fact that camera roll will cause disruptive motion in the video stream, although the roll of the POI is tracked, the actual gimbal roll is not controlled. This assures that the video feed from the visual assistant is always upright, which is desirable for the operator.

V. Experiments

In order to test the controllability of the visual assisting system in as much as possible of its entire configuration space, Fotokite ground station is dismounted from Packbot and placed in the middle of the experimental environment. Since this is only the preliminary work of the whole visual assisting system, the test is conducted in indoor lab environment without obstacles. AprilTag is moved in a random path but covers all four quadrants of the space. All the length units are in AprilTag unit (1 unit = 8.5 cm) and angular units in radians. Fig. 5 shows an example time step of visual servoing interface and its actual pose in world frame. In the left hand side, the small green box represents the desired tag pose and the colored box (blue, red and green lines) is the currently detected tag pose (the large green box is only to visualize 6-DOF tag tracking with depth information). The currently detected pose box should converge to the desired pose box with some disturbances caused by vehicle oscillation. It is not enough that the two squares are co-centered, the four lines of the two squares should also overlap with each other,

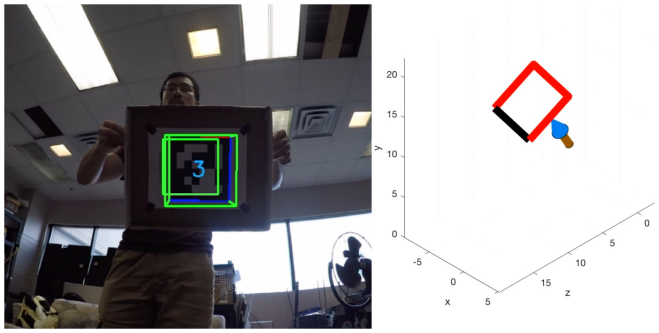


Fig. 5. One Example Time Step of Visual Servoing (Pitch-up). Arrow represents camera's optical axis, and box is AprilTag (with black side down).

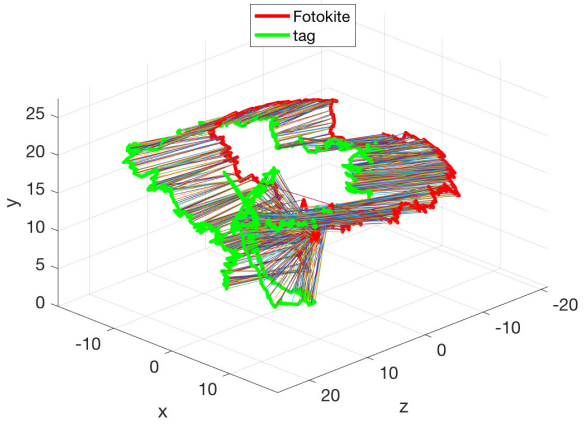


Fig. 6. Trajectory of the POI (Green) and Visual Assistant (Red). Colorful lines connect the origin of the two frames and indicates the constant relative position and orientation from the visual assistant to the POI

indicating that not only POI position, but also depth and orientation are servoed.

A continuos visual assisting trial is displayed in Fig. 6. As we can see, the red trajectory (Fotokite) follows the green trajectory (POI) and Fotokite is maintaining a constant relative position and orientation to the POI.

A closer look into the performance of the same trial is demonstrated in Fig 7. On the left hand side, the profile $\begin{bmatrix} x, & y, & z, & pitch, & yaw, & roll \end{bmatrix}^T$ of the AprilTag and desired vehicle configuration is compared. The profiles are apart by the desired $\mathbf{g}_{\mathbf{t}}^{\mathbf{f}}*$. On the right hand side, the desired and actual vehicle configuration is compared. The two profiles for each state space dimension match with each other, indicating that the visual servoing algorithm is directing Fotokite to the desirable configuration to provide visual assistance.

The error between the desired and actual pose is further investigated in Fig. 8. The mean, root mean square, maximum, and standard deviation of the error for the 3 translations, 3 rotations, total euclidean distance, and rotational norm are summarized in Tab. I. The results indicate that, despite some disturbances caused by UAV's aerial oscillation, sensing inaccuracies and noises, the overal visual servoing process

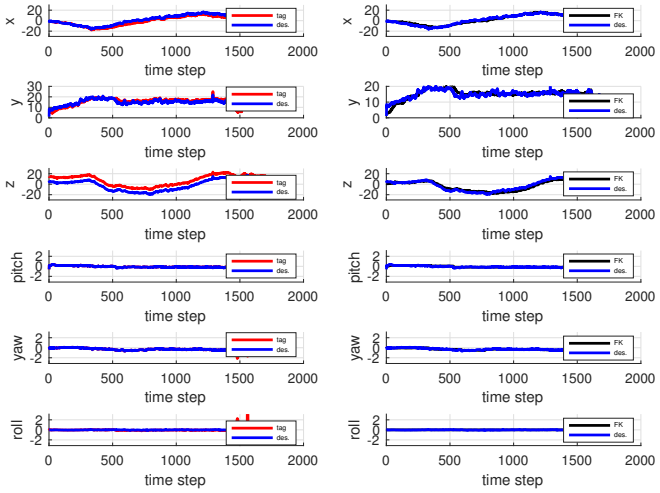


Fig. 7. x, y, z, $pitch$, yaw, and $roll$ of POI, Desired and Actual Visual Assistant Configuration. x, y, and z are in AprilTag units while $pitch$, yaw, and $roll$ are in radians.

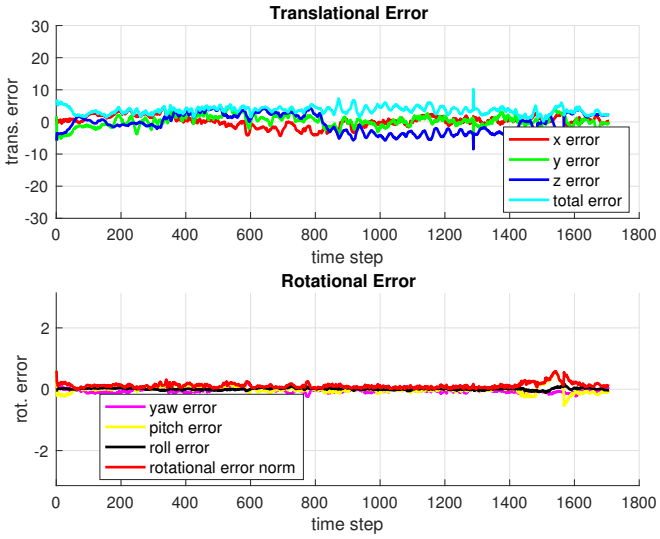


Fig. 8. Error of Translational and Rotational Motion. Translational error is in AprilTag units while rotational error is in radians

can maintain a relatively constant 6-DOF pose from the POI to the visual assistant's camera frame.

VI. Future Work

As mentioned above, this work only lays the groundwork for further teleoperation visual assisting research. Although the experiments conducted in this paper covers as much as possible of the entire configuration space of the visual assistant, the POI has not been installed on the primary robot. The trajectory of the POI mounted on the primary robot may be different. This will possibly lead to different visual assisting behaviors, the performance of which remains to be analyzed.

The environments where teleoperation missions happen are cluttered and occupied with various obstacles. The visual servoing process need to take obstacle avoidance into account. Currently we are installing a LIDAR on Packbot,

TABLE I

MEAN, ROOT MEAN SQUARE, MAXIMUM, AND STANDARD DEVIATION OF SERVOING ERROR

	x	y	z	Euclidean	Yaw	Pitch	Roll	Rotational Norm
Mean	0.2216	0.2473	-0.4885	3.5961	-0.0335	0.0032	0.0080	0.1197
Root Mean Square	1.6998	1.7042	2.8976	3.7669	0.0812	0.1175	0.0324	0.1464
Maximum	4.6605	5.5776	8.4954	10.2147	0.2905	0.5976	0.1935	0.6035
Standard Deviation	1.6858	1.6866	2.8570	1.1217	0.0739	0.1174	0.0314	0.0844

which can provide us with a map of the situation around the primary robot. Then the risk of navigating through obstacles needs to be quantified. So does the visual disturbance to the operator caused by the visual assistant dodging obstacles.

The optimal viewpoint with respect to a certain teleoperation task remains to be determined. We cannot assume looking at the POI straight and keep it in the image center is always the best view angle. Different tasks may have different optimal viewpoints. The optimality of viewpoints may also evolve during the mission. In that case, $\mathbf{g_t^f}*$ is a function of time and introduces more dynamic factors into the system. Perceptual psychology can help with assessing viewpoint quality. Different POIs may be added to different functional units and the visual assistant should switch servo target depending on the task the teleoperator is performing. To sum up, the whole automated visual assisting process should consider a variety of factors, including view point quality, obstacle avoidance and risk assessment, different teleoperation tasks, etc.

VII. CONCLUSION

This paper presents the initial result of the ongoing visual assisting for teleoperation research. Using a tethered UAV, the operator of a ground robot could be visually assisted by a third person view. The extra view point may be beneficial for situational awareness and teleoperation efficiency. Using a fiducial marker as the visual servoing Point of Interest, the assistant UAV is able to track the POI's full state space. Based on a predefined desired viewpoint configuration, the visual servoing algorithm computes the coordinate system transformation and can control the assistant's vehicle and camera pose to maintain a constant 6-DOF relative position and orientation with respect to the POI. Experimental trials have been conducted on physical robot and the performance is quantified and analyzed in terms of control errors in all 6 DOF. The results indicate that the proposed visual servoing approach is able to successfully drive the visual assistant to a moving POI, while maintaining the desired pose for the operator's observation. This approach lays the ground work for future visual assisting research for teleoperation and has the potential to improve teleoperation performance.

ACKNOWLEDGMENT

This work is supported by NSF 1637955, NRI: A Collaborative Visual Assistant for Robot Operations in Unstructured or Confined Environments. The authors would like to thank Mickie Byrd for his help on the visual assistant's ground station server.

REFERENCES

[1] R. Murphy, *Disaster Robotics*. MIT press, 2014.
[2] ——, *Introduction to AI robotics*. MIT press, 2000.
[3] G. J. Agin, *Real time control of a robot with a mobile camera*. SRI International, 1979.
[4] F. Chaumette and S. Hutchinson, "Visual servo control. i. basic approaches," *IEEE Robotics & Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.
[5] ——, "Visual servo control. ii. advanced approaches [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 109–118, 2007.
[6] R. S. Rao, V. Kumar, and C. J. Taylor, "Planning and control of mobile robots in image space from overhead cameras," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2005, pp. 2185–2190.
[7] X. Xiao, E. Cappo, W. Zhen, J. Dai, K. Sun, C. Gong, M. J. Travers, and H. Choset, "Locomotive reduction for snake robots," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3735–3740.
[8] 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*. IEEE, 2011, pp. 4929–4934.
[9] J. Dufek and R. Murphy, "Visual pose estimation of usv from uav to assist drowning victims recovery," in *Safety, Security, and Rescue Robotics (SSRR), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 147–153.
[10] X. Xiao, J. Dufek, T. Woodbury, and R. Murphy, "Uav assisted usv visual navigation for marine mass casualty incident response," in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*. IEEE, 2017.
[11] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE transactions on robotics and automation*, vol. 12, no. 5, pp. 651–670, 1996.
[12] A. C. Sanderson and L. E. Weiss, "Adaptive visual servo control of robots," in *Robot vision*. Springer, 1983, pp. 107–116.
[13] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3400–3407.
[14] R. Murphy, J. Dufek, T. Sarmiento, G. Wilde, X. Xiao, J. Braun, L. Mullen, R. Smith, S. Allred, J. Adams *et al.*, "Two case studies and gaps analysis of flood assessment for emergency management with small unmanned aerial systems," in *Safety, Security, and Rescue Robotics (SSRR), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 54–61.
[15] S. Lupashin and R. D'Andrea, "Stabilization of a flying vehicle on a taut tether using inertial sensing," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 2432–2438.
[16] M. Schulz, F. Augugliaro, R. Ritz, and R. D'Andrea, "High-speed, steady flight with a quadrocopter in a confined environment using a tether," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 2015, pp. 1279–1284.
[17] Fotokite. [Online]. Available: http://fotokite.com/