Best Viewpoints for External Robots or Sensors Assisting Other Robots

Jan Dufek, Xuesu Xiao, and Robin R. Murphy, *Fellow, IEEE*

***Abstract*—This work creates a model of the value of different external viewpoints of a robot performing tasks. The current state of the practice is to use a teleoperated assistant robot to provide a view of a task being performed by a primary robot; however, the choice of viewpoints is ad hoc and does not always lead to improved performance. This research applies a psychomotor approach to develop a model of the relative quality of external viewpoints using Gibsonian affordances. In this approach, viewpoints for the affordances are rated based on the psychomotor behavior of human operators and clustered into manifolds of viewpoints with the equivalent value. The value of 30 viewpoints is quantified in a study with 31 expert robot operators for four affordances (reachability, passability, manipulability, and traversability) using a computer-based simulator of two robots. The adjacent viewpoints with similar values are clustered into ranked manifolds using agglomerative hierarchical clustering. The results show the validity of the affordance-based approach by confirming that there are manifolds of statistically significantly different viewpoint values, viewpoint values are statistically significantly dependent on the affordances, and viewpoint values are independent of a robot. Furthermore, the best manifold for each affordance provides a statistically significant improvement with a large Cohen's d effect size (1.1–2.3) in the performance (improving time by 14%–59% and reducing errors by 87%–100%) and improvement in the performance variation over the worst manifold. This model will enable autonomous selection of the best possible viewpoint and path planning for the assistant robot.**

***Index Terms*—Human–robot interaction, multirobot systems, telerobotics.**

I. Introduction

AN ASSISTANT robot providing a view of a task being performed by a primary robot has emerged as the state of the practice for ground and water robots in homeland security applications, disaster response, and inspection tasks [1]–[6]. Advances in small unmanned aerial systems (UAS), especially tethered UAS, suggest that flying assistant robots will soon supply the needed external visual perspective [7]–[10].

During the 2011 Fukushima Daiichi nuclear power plant accident, teleoperated robots were used in pairs from the beginning of the response to reduce the time it took to accomplish a task [11], [12]. iRobot PackBot unmanned ground vehicles were used to conduct radiation surveys and read dials inside the plant facility, where the assistant PackBot provided camera views of the first robot in order to manipulate door handles, valves, and sensors faster [13].

Since then, the use of two robots to perform a single task has been acknowledged as the best practice for decommissioning tasks. However, the Japanese Atomic Energy Agency has reported through our memorandum of understanding for cooperative research on disaster robotics that operators constantly try to avoid using a robotic visual assistant. The two sets of robot operators find it difficult to coordinate in order to get and maintain the desired view but a single operator becomes frustrated trying to operate both robots.

There are at least two issues with the current state of the practice. First, it increases the cognitive workload on a primary operator by either requiring the primary operator to control two robots or having to coordinate with a secondary operator [12]. Second, it is not guaranteed a human operator will provide ideal viewpoints as viewpoint quality for various tasks is not well understood and humans were shown to pick suboptimal viewpoints [14].

This article addresses the choice of ideal viewpoints by creating a model of the value of different external viewpoints of a robot performing tasks; it is expected, but beyond the scope of this study that the application of the model to robotic visual assistants will likely reduce the cognitive workload on the primary operator. The model will provide an understanding of the utility of different external viewpoints of tasks of the primary robot and can be used as a basis for principled viewpoint selection for a robotic visual assistant. This can ultimately enable autonomous viewpoint selection and path planning for an autonomous robotic visual assistant, therefore, eliminating the need for manual control.

This article is organized as follows. Section II discusses the related work establishing there is no existing model of viewpoint values and showing the importance of psychomotor aspects in viewpoint selection. Section III introduces the affordance-based approach. Section IV details the implementation of a computer-based simulator. Section V presents a human subject study quantifying the value of viewpoints and clustering to create the

Manuscript received December 17, 2020; revised March 31, 2021; accepted May 30, 2021. Date of publication July 13, 2021; date of current version July 19, 2021. This work was supported by NSF 1945105 Best Viewpoints for External Robots or Sensors Assisting Other Robots, and DOE DE-EM0004483 NRI: A Collaborative Visual Assistant for Robot Operations in Unstructured or Confined Environments. This article was recommended by Associate Editor J. A. Adams. *(Corresponding author: Jan Dufek.)*

Jan Dufek and Robin R. Murphy are with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: dufek@tamu.edu; murphy@cse.tamu.edu).

Xuesu Xiao is with the Department of Computer Science, University of Texas at Austin, Austin, TX 78712 USA (e-mail: xiao@cs.utexas.edu).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/THMS.2021.3090765.

Digital Object Identifier 10.1109/THMS.2021.3090765

manifolds. Section VI presents the results showing the validity of the affordance-based approach and a significant improvement in the performance. Section VII discusses the relation to the related work, the reduction in cognitive workload, the ramifications for robotic visual assistants, and the actionable rules for teleoperated robotic visual assistants. Section VIII summarizes the key findings that there are manifolds of different viewpoint values, viewpoint values are dependent on the affordances, and viewpoint values are independent of the robot.

II. Related Work

There is no existing model of viewpoint values leaving robotic visual assistants to rely on ad hoc choices of viewpoints or work envelope models. Woods *et al.* [15]–[24] indicated that improving the ability to comprehend Gibsonian affordances improves teleoperation and external viewpoints improve the ability to comprehend affordances forming an important foundation for the affordance-based approach of this work.

A total of five attributes of an ideal viewpoint were identified and four categories of existing robotic visual assistant implementations were examined with an underlying focus on whether there is an existing model of viewpoint values. There was no existing model of viewpoint values leaving the existing robotic visual assistant implementations to rely on ad hoc choices or on having *a priori* access to, or constructing, 2-D or 3-D models of their work envelope. Robotic visual assistants lacked principles to select ideal viewpoints and no robotic visual assistant implementation considered psychomotor aspects in the viewpoint selection.

There were five attributes of an ideal external viewpoint of action being performed by a robot: the field of view (the area of interest must be in the field of view) [25], visibility/occlusions (the view of the area of interest must be occlusion free) [26], depth of field (the area of interest must be in the depth of field or sharp focus) [27], resolution/zoom (the area of interest must have sufficient resolution in the image so the camera has to be physically close or have to zoom in) [28], and psychomotor aspects (the view must positively affect the human ability to move the robot to accomplish the goal) [22].

There were four categories of robotic visual assistant implementations in the literature all lacking principles to select ideal viewpoints. Static visual assistants [29]–[31] did not move and therefore could not adapt viewpoints to changing pose or actions of the primary robot. Manual visual assistants [32], [33] left the choice of a viewpoint to humans who were previously shown to pick suboptimal viewpoints [14]. Reactive autonomous visual assistants [34]–[36] only reactively tracked and zoomed on the action ignoring the question of what are the best viewpoints. Deliberative autonomous visual assistants [37]–[39] deliberated about certain predefined geometrical criteria while only considering camera configuration attributes of an ideal viewpoint (field of view, visibility/occlusions, depth of field, and resolution/zoom). While camera configuration attributes are necessary preconditions for an ideal viewpoint, no robotic visual assistant studies considered how the viewpoint affects the human teleoperator of the primary robot (psychomotor aspects attribute of an ideal viewpoint) in viewpoint selection.

Psychomotor aspects of an ideal viewpoint were ignored in the existing robotic visual assistant implementations, despite the results by Woods *et al.* [15]–[24] who showed that teleoperation can be improved by improving the ability to comprehend affordances and that an external view improves the ability to comprehend affordances.

Woods *et al.* primarily focused on creating tools to enable humans to manually select external views that supply Gibsonian affordances, which are visual cues that allow humans to directly perceive the possibility of actions independent of the environment or task models [40]. Woods *et al.* contributed two important results forming a theoretical background for this article. They showed that teleoperation can be improved by improving teleoperators' ability to comprehend affordances and they established that an external view improves the ability to comprehend the affordances. This indicates that the value of a viewpoint should depend on the affordances and confirms the benefit of a robotic visual assistant providing an external view.

Despite those contributions, Woods *et al.* relied on human input to select viewpoints and did not evaluate the value of different external viewpoints, they experimentally studied only reachability affordance, they used a simulator that did not reflect realistic robots, and the subjects were not expert robot operators. Their work forms the foundation for the approach in Section III. However, unlike their work, this article creates a model of the value of different external viewpoints (that can ultimately enable a robotic visual assistant to pick a viewpoint without human input), the model is created for four affordances, the simulator used in the experimentation reflects two realistic robots, and the subjects are expert robot operators. Having expert operators using realistic robots prevents confounding the results with subjects struggling to control the robots.

III. Affordance-Based Approach

The approach is to use the concept of Gibsonian affordances [1], where the potential for an action can be directly perceived without knowing intent or models and, thus, is universal to all robots and tasks. In this approach, it is assumed tasks can be decomposed into actions each relying on a single affordance, space around the actions is decomposed into viewpoints, and the viewpoints for the affordances are rated based on teleoperator's psychomotor behavior and clustered into manifolds of viewpoints with the equivalent value (see Fig. 1).

The main postulation of the approach is that a model of viewpoint values can be created using Gibsonian affordances based on psychomotor behavior. This postulation has two central tenets. The first tenet is that the value of a viewpoint depends on the Gibsonian affordance for each action in a task. This tenet is supported by the previous work of Woods *et al.* discussed in Section II. The approach based on Gibsonian affordances has at least two benefits. First, it avoids the need for models seen in the deliberative approach by focusing on the affordances for an action rather than the action itself. Second, research on affordances suggests there are relatively few affordances [41].

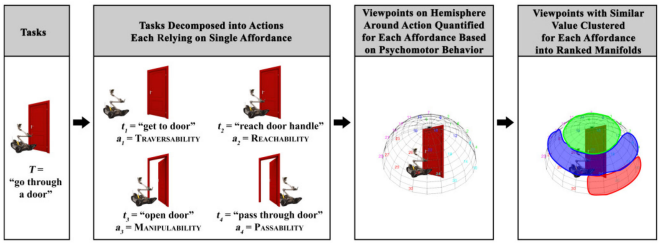

Fig. 1. Overview of the main building blocks of the approach. T, t_i, and a_i denote task, action, and affordance respectively.

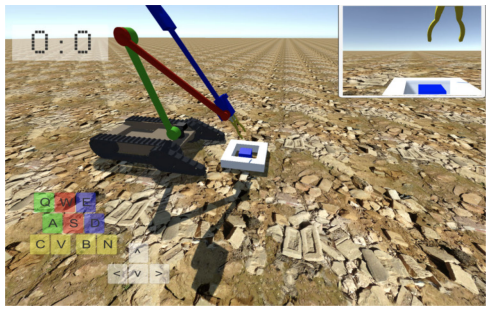

Fig. 2. Reachability affordance: Are the robot and its manipulator in the right pose to reach an object?

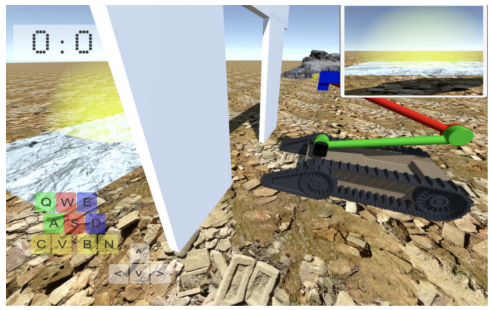

Fig. 3. Passability affordance: Is the robot or its manipulator in the right pose to safely pass through a narrow opening?

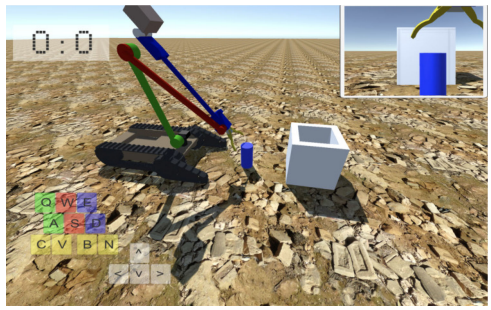

Fig. 4. Manipulability affordance: Is the robot's manipulator in the right pose to manipulate an object?

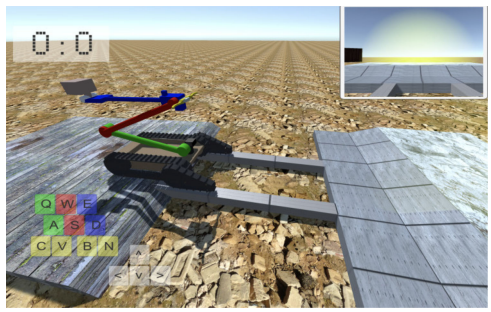

Fig. 5. Traversability affordance: Is the robot in the right pose to safely traverse the environment?

It is, therefore, conceivable every robotic task could be decomposed into a small set of affordances and each affordance would have associated preferred viewpoints. The second tenet is that viewpoints in the space surrounding the action can be rated and adjacent viewpoints with similar ratings can be clustered into manifolds of viewpoints with the equivalent value. The clustering of viewpoints into manifolds has at least three benefits. First, it simplifies navigational reachability. As long as the robotic visual assistant can reach any location within the manifold, it will provide approximately the same value as any other location within the manifold. Second, it aids visual stability. Due to equivalence of viewpoints within the manifold, positioning the robotic visual assistant at the centroid of a manifold will minimize the chance that a potential pose perturbation would significantly change viewpoint quality. Third, it can be used in autonomous planning for a robotic visual assistant to select a manifold and plan a path there while balancing the reward of having a view from that particular manifold with the associated risk of being at that manifold and getting to that manifold [42]–[47] while also considering visual stability.

Based on the related work of Woods *et al.* and our prior experience with 21 disaster deployments, participation in 35 homeland security exercises, and examination of common tasks for robots at Fukushima [12], the development of the model is restricted to four common affordances: reachability (see Fig. 2), passability (see Fig. 3), manipulability (see Fig. 4), and traversability (see Fig. 5). Woods *et al.* additionally discussed climability and drivability affordances, however, those overlap with our traversability (definitions of affordances are not standardized).

Starting with the first building block from Fig. 1, it is assumed every task T can be decomposed into a sequence of actions $t_1, t_2, \ldots, t_n$ where the perception for each action t_i relies on a single affordance, a_i. In reality, actions might rely on a compound affordance, but this work assumes each action relies on its dominant affordance. Then, a task T can be represented by a sequence of action-affordance tuples (t_i, a_i) forming a coarse knowledge representation of the task.

Space around the action can be decomposed into viewpoints that are assumed to be lying on a hemisphere of a fixed radius (see Fig. 6). A viewpoint is represented using a spherical coordinate system as $v = (r, \theta, \varphi)$ and the optical axis is along the radius r. While the values of r can vary in practice, an assumption for this work is that a hemisphere with a fixed radius r serves as the idealized workspace envelope for the assistant.

A viewpoint v will have a value $|v|$ based on how well a teleoperator can perform the action from that viewpoint. The value is composed of the time to complete the action and the number of errors.

Adjacent viewpoints v with similar value $|v|$ will form a continuous volume, or manifold, M. Within a manifold, each

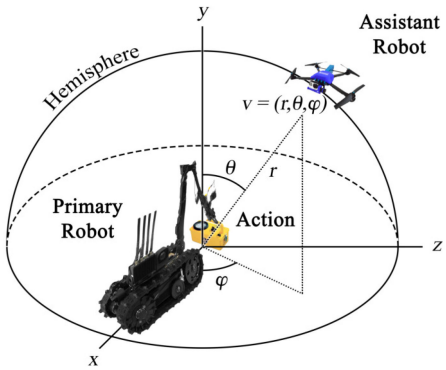

Fig. 6. Hemisphere with a fixed radius r centered in the action serves as an idealized work envelope for the robotic visual assistant.

viewpoint is equally good. The entire space will be divided into ranked manifolds.

The model of viewpoint values will be extracted in two steps. First, the value of viewpoints $|v|$ for the four affordances will be quantified in a human subject study using a computer-based simulation. Second, adjacent viewpoints of similar value will be clustered into manifolds of viewpoints with the equivalent value.

IV. Simulator Implementation

A computer-based simulator was created to enable the quantification of the value of viewpoints by remotely (over the web) measuring the performance of expert robot operators controlling one of two robots (iRobot PackBot or QinetiQ TALON) in four tasks corresponding to the four affordances from different external viewpoints. Using expert robot operators already proficient with the robots reduces the chance of confounding the results with subjects' varying familiarity with the robots, varying difficulty in controlling the robots, and varying time needed to train on the robots. It also reduces learning effects as subjects unfamiliar with the robots might gradually learn how to control the robots during the experiment. Those two specific models of robots were selected because they are the two most common explosive ordnance disposal robots making it easier to find subjects proficient with at least one of them. The use of computer-based simulation is justified based on previous work of Woods *et al.* (see Section II) who showed computer-based simulation is suitable to measure the teleoperators' ability to comprehend Gibsonian affordances. The simulator was implemented in C# using the Unity engine and runs on Amazon Web Services (AWS) infrastructure. The AWS S3 supports a front-end website with the Unity simulation interface while AWS EC2 runs a back-end responsible for receiving and storing the data. When running the simulation, subjects can see a large external view of the task from a specific viewpoint, a small fixed view from a forward-looking onboard camera of the primary robot, a color-coded keyboard legend corresponding to the color-coding of the primary's robot arm (this is necessary because a keyboard is not a typical mode of control of those robots), and a clock to constantly remind them they are being timed (as seen in Figs. 2–5). When a subject makes an error, the error location is highlighted in red and an error sound is played to make the subject aware of the error.

V. Experimentation

The experimentation is done by quantifying the value of viewpoints in a human subject study and then clustering the viewpoints into ranked manifolds. The value of 30 viewpoints is quantified in a 31 person human subject study for 4 Gibsonian affordances (reachability, passability, manipulability, and traversability) using a computer-based simulator. The data from the human subject study are then used to rate the viewpoints and cluster adjacent viewpoints with similar value into manifolds of viewpoints with the equivalent value using agglomerative hierarchical clustering.

A. Quantifying Viewpoints in Human Subject Study

A 31 person human subjects study was designed with a goal to sufficiently sample human performance for 30 viewpoints v_i, where $i = 1, \ldots, 30$, to quantify the value of viewpoints $|v_i^a|$ for each of the four affordances a so that spatial clusters (manifolds) can be learned. The subjects perform four tasks corresponding to the four affordances from varying external viewpoints while their performance is measured in terms of time and number of errors to quantify the corresponding viewpoint value.

The subjects are 31 (based on power analysis) male expert robot operators of age ranging from 23 to 46 years ($M = 31.5$, SD $= 5.9$) experienced with either PackBot or TALON robots. The subjects use their own computer to connect to a remote computer-based simulator via a web browser. The subjects choose either PackBot or TALON robots based on their experience (10 subjects chose PackBot and 21 chose TALON).

The subjects perform four kinds of tasks each associated with one of the four affordances. For reachability, the task is to touch the blue cube using the gripper without hitting the neighboring blocks (see Fig. 2). For passability, the task is to pass through the opening in the walls and take caution to not hit the walls (see Fig. 3). For manipulability, the task is to pick up the blue cylinder and drop it in the bin without hitting the bin with the gripper (see Fig. 4). For traversability, the task is to cross the ridge and reach the other side without falling on the ground (see Fig. 5).

The independent variable is the position of the external viewpoint provided to the subject. A total of 30 possible viewpoints, v_i where $i = 1, \ldots, 30$, are equidistantly dispersed on a hemisphere with a fixed radius of $r = 1.5$ m centered at the task location at (0,0,0), as illustrated in Fig. 7. The distance between viewpoints is approximately 0.7 m. Those 30 viewpoints are divided into 5 groups (6 viewpoints per group) based on their relative position to the task location: left, right, front, back, and top. Each subject performs each of the 4 tasks from each of the 5 viewpoints groups (20 rounds total). The particular viewpoint from each group is always selected randomly. The order of the tasks and viewpoint groups is randomized for each subject to reduce the order effect. The five viewpoint groups are used solely to help the samples to be uniformly distributed across viewpoints and the specific choice of groups does not have an

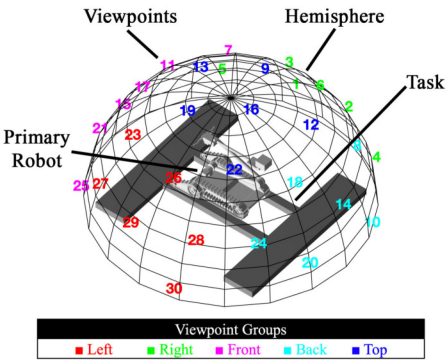

Fig. 7. Total of 30 possible viewpoints are equidistantly dispersed on a hemisphere centered at the task and divided into five groups. The figure is in scale.

influence on the final results. An alternative approach would be to manually distribute the viewpoints to the subjects in a way that each viewpoint has the same number of samples.

There are two dependent variables both indicating the subject's performance: the time to complete the task and the number of errors. Those two measures were the most common in the reviewed robotic visual assistant studies. For the reachability and manipulability tasks, the number of errors is the number of manipulator collisions. For the passability task, the number of errors is the number of robot collisions. For the traversability task, the number of errors is the number of falls of the robot.

The metric indicating the quality of a viewpoint is the subject's performance computed as

$$ {}_jP_i^a = -w_t\left(\widetilde{{}_jt_i^a}\right) - w_e\left(\widetilde{{}_je_i^a}\right) \tag{1} $$

where

$$ \widetilde{{}_jt_i^a} = \frac{{}_jt_i^a - \text{mean}_{a',i'}\left({}_jt_{i'}^{a'}\right)}{\text{std}_{a',i'}\left({}_jt_{i'}^{a'}\right)} \tag{2} $$

$$ \widetilde{{}_je_i^a} = \frac{{}_je_i^a - \text{mean}_{a',i'}\left({}_je_{i'}^{a'}\right)}{\text{std}_{a',i'}\left({}_je_{i'}^{a'}\right)}. \tag{3} $$

j is a subject index, i is a viewpoint index, a is an affordance index, ${}_jP_i^a$ denotes the performance of subject j for affordance a from viewpoint v_i, ${}_jt_i^a$ is the time subject j took to complete the task associated with affordance a from viewpoint v_i, ${}_je_i^a$ is the number of errors subject j made when performing the task associated with affordance a from viewpoint v_i, $\widetilde{{}_jt_i^a}$ and $\widetilde{{}_je_i^a}$ are ${}_jt_i^a$ and ${}_je_i^a$ normalized across all samples for subject j, and w_t and w_e are the weights of the time term and error term of the performance, respectively. The formula uses time and errors that are normalized for each individual subject to reduce the effects of the variation in overall performance between individual subjects. The weighted sum in this formula is multiplied by -1 so that the performance is more intuitive to interpret. Without this adjustment, the lower performance would be better because less time and fewer errors are better, however, the lower performance being better is counterintuitive. The w_t and w_e weights are set to 0.4 and 0.6, respectively, for this experiment. This weights the errors slightly higher than completion time to penalize completing a task faster at the expense of more errors.

The study results in a set of performance samples where one performance sample ${}_jP_i^a$ represents a performance of subject j at viewpoint v_i for affordance a. A performance sample is rejected as an outlier if the corresponding ${}_jt_i^a$ value is more than three scaled median absolute deviations away from the $\text{med}_{j'j'}t_i^a$. An example of an outlier would be a subject getting distracted in the middle of a task (e.g., answering a phone call) causing their time to complete the task to be higher than it would have been.

The value of a viewpoint v_i for affordance a is defined as

$$ |v_i^a| = w_m\,\text{mean}_j\left({}_jP_i^a\right) - w_d\,\text{std}_j\left({}_jP_i^a\right) \tag{4} $$

where w_m and w_d are the weights of the mean and standard deviation term of the viewpoint value, respectively. While the viewpoint value should be primarily indicated by the mean of the corresponding performance samples, the standard deviation term is introduced to also make the viewpoint value inversely proportional to the standard deviation of the corresponding performance samples (higher standard deviation indicates higher unpredictability of the performance). The w_m and w_d weights are set to 0.9 and 0.1, respectively, for this experiment. This makes the mean term the dominant indicator of the viewpoint value.

B. Learning Manifolds by Clustering Viewpoints

Agglomerative hierarchical cluster analysis with average linkages [48] is used to generate manifolds. Pairwise dissimilarity is computed using the combination of the orthodromic distance and the difference between the normalized viewpoint values, and used to construct a hierarchical cluster tree using the unweighted pair group method with arithmetic mean linkage. The number of manifolds is determined by maximizing the Calinski–Harabasz criterion [49]. The value of a manifold is computed as a combination of the mean and standard deviation of the values of the member viewpoints. To be able to compare two manifolds in terms of time and errors, a metric comparing the time and errors using the intersection of the subjects in the two manifolds is introduced.

The input is 30 sample points for each affordance a, where one sample point is

$$ s_i^a = (\theta_i, \varphi_i, \widetilde{|v_i^a|}) \tag{5} $$

where θ_i and φ_i are the polar angle and azimuthal angle of ith viewpoint v_i in the spherical coordinate system and $\widetilde{|v_i^a|} = \frac{|v_i^a| - \text{mean}_{i'=1,\ldots,30}|v_{i'}^a|}{\text{std}_{i'=1,\ldots,30}|v_{i'}^a|}$ is the normalized value of ith viewpoint, v_i, for affordance a.

The pairwise dissimilarity between all 30 sample points s_i^a, where $i = 1, \ldots, 30$, for affordance a is computed using the combination of the orthodromic distance of the viewpoints on the hemisphere and the difference between the normalized viewpoint values resulting in 435 dissimilarities

$$ d_{s_i^a s_j^a} = \sqrt{\left(d_{s_i^a s_j^a}^{(o)}\right)^2 + \left(d_{s_i^a s_j^a}^{(p)}\right)^2} \tag{6} $$

where $s_i^a = (\theta_i, \varphi_i, \widetilde{|v_i^a|})$ and $s_j^a = (\theta_j, \varphi_j, \widetilde{|v_j^a|})$ are two sample points, $d^{(o)}_{s_i^a s_j^a}$ is the orthodromic distance between the two sample points, and $d^{(p)}_{s_i^a s_j^a}$ is the value distance between the two sample points. The orthodromic distance is the great-circle distance of the two associated viewpoints on the hemisphere defined as

$$d^{(o)}_{s_i^a s_j^a} = 2r \operatorname{atan2}\left(\sqrt{\xi}, \sqrt{1-\xi}\right) \tag{7}$$

where

$$\xi = \sin^2 \frac{\theta_j - \theta_i}{2} + \cos\theta_i \cos\theta_j \sin^2 \frac{\varphi_j - \varphi_i}{2}. \tag{8}$$

The value distance is the difference between the normalized values of the two associated viewpoints defined as

$$d^{(p)}_{s_i^a s_j^a} = \left| \widetilde{|v_i^a|} - \widetilde{|v_j^a|} \right|. \tag{9}$$

The sample points are grouped into a binary hierarchical cluster tree using the unweighted pair group method with arithmetic mean linkage [48]. The linkage between clusters (manifolds) M_k^a and M_l^a for affordance a is the average linkage defined as

$$d_{M_k^a M_l^a} = \frac{1}{N_{M_k^a} N_{M_l^a}} \sum_{s_i^a \in M_k^a} \sum_{s_j^a \in M_l^a} d_{s_i^a s_j^a} \tag{10}$$

where $N_{M_k^a}$ is the number of sample points in kth cluster (manifold), M_k^a, for affordance a.

The number of manifolds N_m^a for affordance a is determined by maximizing the Calinski–Harabasz criterion [49]. This criterion is used based on a premise that well-defined clusters have a large between-cluster variance and a small within-cluster variance. For this work, the number of manifolds is limited to 10 to prevent forming too many small manifolds.

The value of a manifold M_k^a is defined as

$$|M_k^a| = w_m \operatorname{mean}_{v_i \in M_k^a} |v_i^a| - w_d \operatorname{std}_{v_i \in M_k^a} |v_i^a| \tag{11}$$

where $v_i \in M_k^a \iff s_i^a \in M_k^a$.

To be able to compare two manifolds in terms of the non-normalized time and number of errors, a metric quantifying a relative improvement in the time and number of errors between two manifolds is introduced. This metric measures the improvement only on the intersection of the subjects in the two manifolds to prevent biasing the relative improvement with the variation in overall speed among individual subjects and is defined as

$$I^{(t)}_{M_k^a M_l^a} = \frac{{}_{M_k^a}t^a_{M_l^a} - {}_{M_l^a}t^a_{M_k^a}}{{}_{M_k^a}t^a_{M_l^a}} \tag{12}$$

where $I^{(t)}_{M_k^a M_l^a}$ is the relative improvement in time t of manifold M_k^a over manifold M_l^a, ${}_{M_l^a}t^a_{M_k^a} = \operatorname{mean}_{j \in \mathcal{S}(M_k^a) \cap \mathcal{S}(M_l^a)} {}_j t^a_{M_k^a}$ is the average time to complete the task associated with affordance a from manifold M_k^a measured by only taking subjects that have at least one sample in both M_k^a and M_l^a manifolds, ${}_j t^a_{M_k^a} = \operatorname{mean}_{v_i \in M_k^a} {}_j t_i^a$ is the average time subject j took to complete the task associated with affordance a from manifold M_k^a, and $\mathcal{S}(M_k^a)$ is the set of subjects that have at least one sample in manifold M_k^a. The relative improvement in the errors $I^{(e)}_{M_k^a M_l^a}$ is measured analogically.

VI. Results

The results show the validity of the affordance-based approach by confirming there are manifolds of statistically significantly different viewpoint values, viewpoint values depend on the affordances, and viewpoint values are independent of a robot. The best manifold for each affordance provides a statistically significant improvement with a large Cohen's d effect size (1.1–2.3) in the performance (improving time by 14%–59% and reducing errors by 87%–100%) and improvement in the performance variation over the worst manifold. All statistical testing is on significance level $\alpha = 0.05$.

A. Validity of Affordance-Based Approach

The results support the two central tenets of the approach by confirming that there are manifolds of statistically significantly different viewpoint values, viewpoint values are statistically significantly dependent on the affordances, and viewpoint values are independent of a robot.

There are manifolds of statistically significantly different viewpoint values. Not all views are equal, and some manifolds provide statistically significantly better views than others. This is tested using an unbalanced one-way analysis of variance (ANOVA) test for each affordance testing that not all $P^a_{M_k^a}$ for $k = 1, \ldots, N_m^a$ for a specific affordance a are equal, where $P^a_{M_k^a} = \operatorname{mean}_{v_i \in M_k^a} ({}_j P_i^a)$ is the mean of all performance samples in manifold M_k^a for affordance a. This is confirmed for all affordances a based on F-statistics and p-values listed in Table I.

The viewpoint values are statistically significantly dependent on the affordances. This is tested using an unbalanced two-way ANOVA test for interaction effects testing whether there is an interaction between affordance factor a and viewpoint factor i for response variable $\widetilde{P^a_{v_i}}$, where $\widetilde{P^a_{v_i}} = \operatorname{mean}_j (\widetilde{{}_j P_i^a})$ is the mean normalized performance for viewpoint v_i for affordance a and $\widetilde{{}_j P_i^a} = \frac{{}_j P_i^a - \operatorname{mean}_{i',j'} ({}_{j'} P^a_{i'})}{\operatorname{std}_{i',j'} ({}_{j'} P^a_{i'})}$ is the performance ${}_j P_i^a$ normalized within the affordance a (normalization is necessary because different affordances have different scales of the performance). The interaction is confirmed based on F-statistic $F(86, 457) = 1.8361$ and p-value 4.0652×10^{-5}.

The viewpoint values are independent of the robot. The selected robot does not have a statistically significant influence on viewpoint values (i.e., the viewpoint values are very similar for both robots). This is tested using an unbalanced two-way ANOVA test for interaction effects testing whether there is an interaction between robot factor ρ and viewpoint factor i for response variable $\widetilde{{}^\rho P_{v_i}}$, where $\widetilde{{}^\rho P_{v_i}} = \operatorname{mean}_{j,a} (\widetilde{{}^\rho_j P_i^a})$ is the mean normalized performance for viewpoint v_i for robot ρ, $\widetilde{{}^\rho_j P_i^a} = \frac{{}^\rho_j P_i^a - \operatorname{mean}_{i',j',\rho'} ({}^{\rho'}_{j'} P^a_{i'})}{\operatorname{std}_{i',j',\rho'} ({}^{\rho'}_{j'} P^a_{i'})}$ is the ${}^\rho_j P_i^a$ normalized within the affordance a (normalization is necessary because different affordances have different scales of the performance), and ${}^\rho_j P_i^a$ is the performance of subject j for affordance a and robot ρ from viewpoint v_i. The interaction is not confirmed based on F-statistic $F(29, 516) = 1.1507$ and p-value 0.27078. Since the

TABLE I
THERE ARE MANIFOLDS OF STATISTICALLY SIGNIFICANTLY DIFFERENT VIEWPOINT VALUES

Affordance (a) ▶	REACHABILITY		PASSABILITY						MANIPULABILITY										TRAVERSABILITY						
Number of Manifolds (N_m^a)	2		6						10										7						
Manifold (M_k^a) ▶	M_1^{R}	M_2^{R}	M_1^{P}	M_2^{P}	M_3^{P}	M_4^{P}	M_5^{P}	M_6^{P}	M_1^{M}	M_2^{M}	M_3^{M}	M_4^{M}	M_5^{M}	M_6^{M}	M_7^{M}	M_8^{M}	M_9^{M}	M_{10}^{M}	M_1^{T}	M_2^{T}	M_3^{T}	M_4^{T}	M_5^{T}	M_6^{T}	M_7^{T}
Number of Samples $\left(N_{M_k^a}^{(P)}\right)$	109	33	34	47	18	5	25	12	5	15	24	18	29	24	14	14	6	1	29	17	28	39	15	14	1
Performance Mean $\left(P_{M_k^a}^a\right)$	0.52	0.09	0.5	0.46	0.35	−0.01	−0.25	−0.5	0.27	0.09	0.05	0.01	−0.45	−0.55	−0.51	−0.67	−1.5	−1.8	0.26	0.27	−0.13	−0.33	−0.71	−0.97	−2.6
F-**statistic**	$F(1,140)$ $= 30.3359$		$F(5,135) = 19.1088$						$F(9,140) = 5.4869$										$F(6,136) = 7.4848$						
p-**value**	1.6872×10^{-7}		2.4179×10^{-14}						1.8816×10^{-6}										6.083×10^{-7}						

TABLE II
REACHABILITY: 14% TIME AND 87% ERROR IMPROVEMENT

Affordance (a) ▶	REACHABILITY		
Number of Manifolds (N_m^a)	2		
Manifold (M_k^a) ▶	**Best** (M_1^{R})	**Worst** (M_2^{R})	**Improvement** $\left(I_{M_1^{\mathrm{R}} M_2^{\mathrm{R}}}\right)$
Manifold Value $\left(\lvert M_k^a \rvert\right)$	0.387	−0.032	-
Time	${}_{M_2^{\mathrm{R}}}t_{M_1^{\mathrm{R}}}^{\mathrm{R}} = 21.11\,\mathrm{s}$	${}_{M_1^{\mathrm{R}}}t_{M_2^{\mathrm{R}}}^{\mathrm{R}} = 24.58\,\mathrm{s}$	**14%**
Errors	${}_{M_2^{\mathrm{R}}}e_{M_1^{\mathrm{R}}}^{\mathrm{R}} = 0.08$	${}_{M_1^{\mathrm{R}}}e_{M_2^{\mathrm{R}}}^{\mathrm{R}} = 0.613$	**87%**

TABLE III
PASSABILITY: 23% TIME AND 100% ERROR IMPROVEMENT

Affordance (a) ▶	PASSABILITY		
Number of Manifolds (N_m^a)	6		
Manifold (M_k^a) ▶	**Best** (M_1^{P})	**Worst** (M_6^{P})	**Improvement** $\left(I_{M_1^{\mathrm{P}} M_6^{\mathrm{P}}}\right)$
Manifold Value $\left(\lvert M_k^a \rvert\right)$	0.376	−0.47	-
Time	${}_{M_6^{\mathrm{P}}}t_{M_1^{\mathrm{P}}}^{\mathrm{P}} = 20.53\,\mathrm{s}$	${}_{M_1^{\mathrm{P}}}t_{M_6^{\mathrm{P}}}^{\mathrm{P}} = 26.74\,\mathrm{s}$	**23%**
Errors	${}_{M_6^{\mathrm{P}}}e_{M_1^{\mathrm{P}}}^{\mathrm{P}} = 0$	${}_{M_1^{\mathrm{P}}}e_{M_6^{\mathrm{P}}}^{\mathrm{P}} = 1.2$	**100%**

TABLE IV
MANIPULABILITY: 39% TIME AND 100% ERROR IMPROVEMENT

Affordance (a) ▶	MANIPULABILITY		
Number of Manifolds (N_m^a)	10		
Manifold (M_k^a) ▶	**Best** (M_1^{M})	**Second-Worst** (M_9^{M})	**Improvement** $\left(I_{M_1^{\mathrm{M}} M_9^{\mathrm{M}}}\right)$
Manifold Value $\left(\lvert M_k^a \rvert\right)$	0.212	−1.379	-
Time	${}_{M_9^{\mathrm{M}}}t_{M_1^{\mathrm{M}}}^{\mathrm{M}} = 29.57\,\mathrm{s}$	${}_{M_1^{\mathrm{M}}}t_{M_9^{\mathrm{M}}}^{\mathrm{M}} = 48.84\,\mathrm{s}$	**39%**
Errors	${}_{M_9^{\mathrm{M}}}e_{M_1^{\mathrm{M}}}^{\mathrm{M}} = 0$	${}_{M_1^{\mathrm{M}}}e_{M_9^{\mathrm{M}}}^{\mathrm{M}} = 0.5$	**100%**

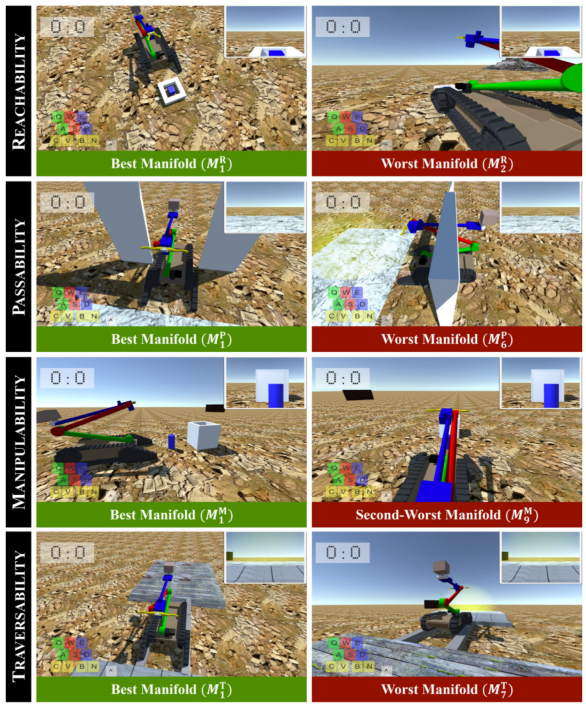

Fig. 8. View from the centroid of the best manifold as compared to the worst manifold.

TABLE V
TRAVERSABILITY: 59% TIME AND 100% ERROR IMPROVEMENT

Affordance (a) ▶	TRAVERSABILITY		
Number of Manifolds (N_m^a)	7		
Manifold (M_k^a) ▶	**Best** (M_1^{T})	**Worst** (M_7^{T})	**Improvement** $\left(I_{M_1^{\mathrm{T}} M_7^{\mathrm{T}}}\right)$
Manifold Value $\left(\lvert M_k^a \rvert\right)$	0.197	−2.108	-
Time	${}_{M_7^{\mathrm{T}}}t_{M_1^{\mathrm{T}}}^{\mathrm{T}} = 23.47\,\mathrm{s}$	${}_{M_1^{\mathrm{T}}}t_{M_7^{\mathrm{T}}}^{\mathrm{T}} = 57.02\,\mathrm{s}$	**59%**
Errors	${}_{M_7^{\mathrm{T}}}e_{M_1^{\mathrm{T}}}^{\mathrm{T}} = 0$	${}_{M_1^{\mathrm{T}}}e_{M_7^{\mathrm{T}}}^{\mathrm{T}} = 5$	**100%**

interaction term of the ANOVA test is nonsignificant, the interaction effect is either very small and statistically nonsignificant or does not exist.

B. *Statistically Significant Improvement in Performance*

The results show there is a statistically significant improvement with a large Cohen's d effect size (1.1–2.3) between the best and worst manifold for each affordance improving time by 14% to 59% and reducing errors by 87% to 100%. The best manifolds also provide an improvement in the performance variation over the worst manifolds.

The best manifold for each affordance improves time by 14% to 59% and reduces errors by 87% to 100% over the worst manifold. Fig. 8 shows for each affordance a view from the best manifold as compared to the worst manifold. Tables III –V provide a quantitative comparison between the best and worst manifold for each affordance (for manipulability, the second-worst manifold is used for this comparison as there are no subjects in the intersection of the worst and best manifold). Figs. 9 –12 show the visualization of the manifolds for each affordance.

The best manifold is statistically significantly better than the worst manifold for each affordance with a large Cohen's d effect size (1.1–2.3). This is tested using one-tailed two-sample t-test (left-tailed) testing for each affordance a that $P_{M_{w(a)}^a}^a < P_{M_{b(a)}^a}^a$, where $M_{w(a)}^a$ and $M_{b(a)}^a$ are the worst and best manifolds for affordance a, respectively, $w(a) = \arg\min_{k=1,\ldots,N_m^a} |M_k^a|$, and $b(a) = \arg\max_{k=1,\ldots,N_m^a} |M_k^a|$. The worst manifold for manipulability (M_{10}^{M}) and traversability (M_7^{T}) was replaced by the second-worst manifold, M_9^{M} and M_6^{T}, respectively, because the worst manifolds each have only one sample. The hypothesis is confirmed for all affordances a based on t-statistics and p-values listed in Table VI. The table also lists Cohen's d effect

TABLE VI
BEST MANIFOLDS STATISTICALLY SIGNIFICANTLY BETTER THAN WORST WITH LARGE COHEN'S d EFFECT SIZE

Affordance (a) ►	**REACHABILITY**		**PASSABILITY**		**MANIPULABILITY**		**TRAVERSABILITY**	
Number of Manifolds (N_m^a)	2		6		10		7	
Manifold (M_k^a) ►	**Best** (M_1^{R})	**Worst** (M_2^{R})	**Best** (M_1^{P})	**Worst** (M_6^{P})	**Best** (M_1^{M})	**Worst** $(M_9^{\mathrm{M}})^*$	**Best** (M_1^{T})	**Worst** $(M_6^{\mathrm{T}})^*$
Number of Samples $\left(N_{M_k^a}^{(P)}\right)$	109	33	34	12	5	6	29	14
Performance Mean $\left(P_{M_k^a}^a\right)$	0.51582	0.093026	0.50487	−0.49656	0.27369	−1.5455	0.25833	−0.96995
Performance Std. Dev. $\left(\sigma\left(P_{M_k^a}^a\right)\right)$	0.25781	0.65474	0.13714	0.84357	0.097376	1.4138	0.2644	1.2145
t**-statistic**	−3.6254		−4.0933		−3.1429		−3.7416	
p**-value**	0.00045367		0.0008565		0.012595		0.0011469	
Cohen's d **Effect Size** $(\mathcal{D}^a)$	1.0943		2.2854		1.7231		1.7109	

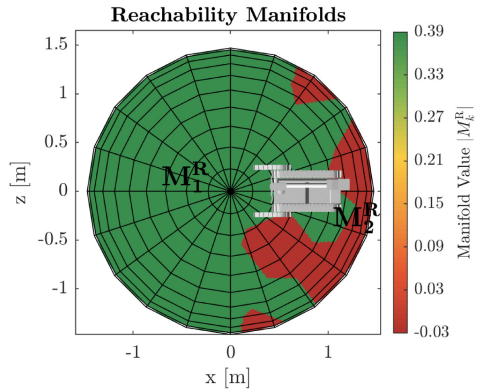

Fig. 9. Two manifolds for reachability affordance. The figure shows a top-down view of the hemisphere with the two manifolds (M_1^{R} and M_2^{R}) color-coded by the corresponding manifold value ($|M_1^{\mathrm{R}}|$ and $|M_2^{\mathrm{R}}|$). The manifold labels are placed in the centroid of the corresponding manifold. The manifolds are indexed in ascending order from the best to the worst. The figure is in scale including the size and pose of the primary robot (colored gray).

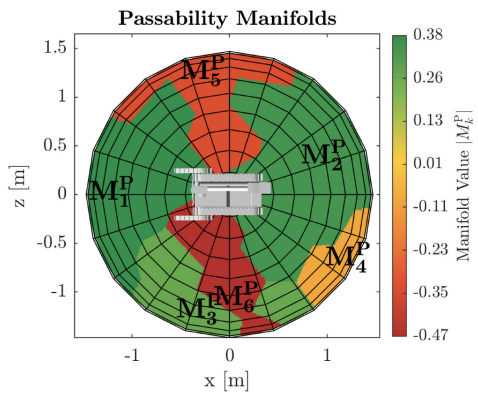

Fig. 10. Six manifolds for passability affordance. The figure shows a top-down view of the hemisphere with the six manifolds ($M_1^{\mathrm{P}}, \ldots, M_6^{\mathrm{P}}$) color-coded by the corresponding manifold value ($|M_1^{\mathrm{P}}|, \ldots, |M_6^{\mathrm{P}}|$).

size, number of performance samples $N_{M_k^a}^{(P)}$, and the standard deviation of the performance $\sigma(P_{M_k^a}^a) = \mathrm{std}_{v_i \in M_k^a}\,({}_jP_i^a)$ for manifold M_k^a.

The best manifolds provide an improvement in the performance variation over the worst manifolds. This is because good viewpoints are consistently good across subjects but bad viewpoints have a large variation in the performance (time and errors). This means that having a view from a good manifold leads to a good predictable performance while having a view from a bad

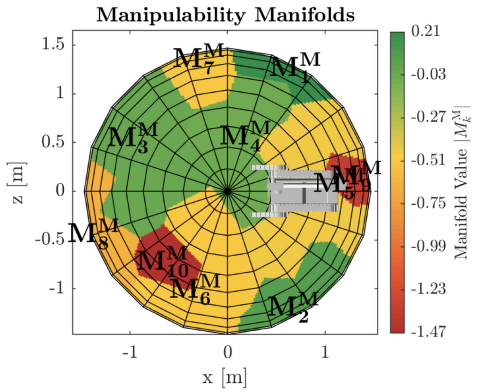

Fig. 11. Ten manifolds for manipulability affordance. The figure shows a top-down view of the hemisphere with the ten manifolds ($M_1^{\mathrm{M}}, \ldots, M_{10}^{\mathrm{M}}$) color-coded by the corresponding manifold value ($|M_1^{\mathrm{M}}|, \ldots, |M_{10}^{\mathrm{M}}|$).

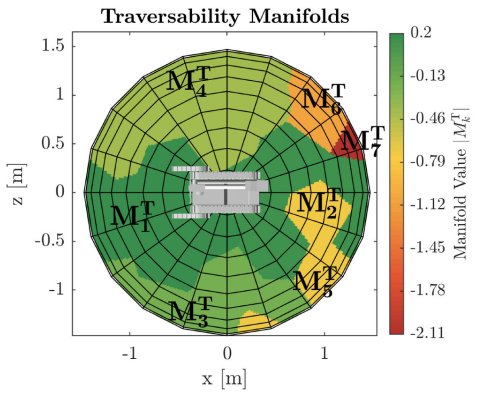

Fig. 12. Seven manifolds for traversability affordance. The figure shows a top-down view of the hemisphere with the seven manifolds ($M_1^{\mathrm{T}}, \ldots, M_7^{\mathrm{T}}$) color-coded by the corresponding manifold value ($|M_1^{\mathrm{T}}|, \ldots, |M_7^{\mathrm{T}}|$).

manifold not only leads to a bad performance but also leads to unpredictability in what might go wrong and how much. Fig. 13 illustrates this on passability affordance.

VII. DISCUSSION

The results are consistent with Woods *et al.* confirming that an external viewpoint improves teleoperation, and further showing there are manifolds of viewpoints with some manifolds being significantly better than others. While beyond the scope, it is expected the proposed model will likely reduce the cognitive workload by eliminating the need to manually control the

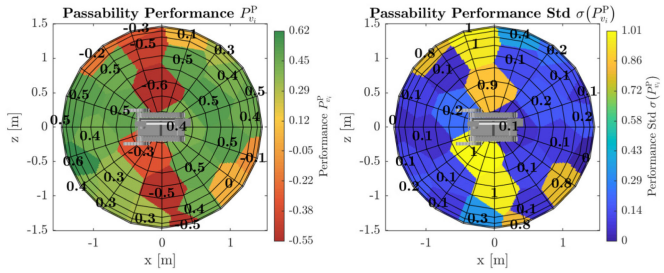

Fig. 13. Viewpoints with a low mean performance (red) also have a high standard deviation of the performance (yellow). The performance (left) and the standard deviation of the performance (right) is defined as $P^a_{v_i} = \mathrm{mean}_j\,({}_jP^a_i)$ and $\sigma(P^a_{v_i}) = \mathrm{std}_j\,({}_jP^a_i)$, respectively.

robotic visual assistant and improving the ability to comprehend affordances. The results have implications for robotic visual assistants using the manifolds both in terms of visual stability and tracking and can be extracted into actionable rules. The manifolds for all affordances except manipulability are not very sensitive to the weights used in the viewpoint value and manifold value definitions, it is, therefore, not necessary to adjust them.

A. Relation to Related Work

The results are consistent with Woods *et al.*, who showed an external view improves the ability to comprehend affordances making teleoperation easier. This work goes further by showing that not all external views are equal and that there are different regions of viewpoints (manifolds) where good manifolds significantly improve time and reduce errors over bad manifolds. This work is also in agreement with the position of Woods *et al.* that there is no best viewpoint for a task. The viewpoint values in the proposed model depend on the affordances, and therefore, the viewpoint would change over time as a task may consist of a series of affordances. For example, a task "go through a door" may consist of four affordances—traversability to get to the door, reachability to reach the door handle, manipulability to open the door, and passability to pass through the door (illustrated in Fig. 1). In this example, the best viewpoint would change four times during a single task.

B. Reduction in Cognitive Workload

While beyond the scope of this article, it is expected the model will likely reduce the cognitive workload on the primary robot operator. The model will enable to make the robotic visual assistant autonomous eliminating the need for the primary robot operator to manually control the robotic visual assistant or to coordinate with a secondary operator [7]–[10], [42]–[47]. The model will also allow the robotic visual assistant to select a viewpoint for each action that enables direct apprehension of the affordance for that action reducing the need for high-workload deliberative reasoning about the properties of the scene that would be required if the affordance for the action could not be directly perceived, as shown by Morison [23].

C. Ramifications for Robotic Visual Assistants

The results have implications for robotic visual assistants using the manifolds both in terms of visual stability and tracking. The best manifolds for all the affordances except manipulability have large areas (77%, 23%, 7%, and 20% of the hemisphere surface for reachability, passability, manipulability, traversability, respectively) suggesting that positioning a robotic visual assistant in the best manifold centroid will result in good visual stability (potential perturbations in the pose of a robotic visual assistant will not significantly change the view quality). The best manifolds for reachability (M^{R}_1) and manipulability (M^{M}_1 and M^{M}_2) are shifted toward the robot indicating those affordances are object–robot centric and suggesting the necessity to track both the object and the end effector as can be seen in Figs. 9 and 11, respectively. The best manifolds for passability (M^{P}_1 and M^{P}_2) and traversability (M^{T}_1 and M^{T}_2) are elongated along approach and departure directions indicating those affordances are robot-centric and suggesting the necessity to track the entire action (movement) of the robot as can be seen in Figs. 10 and 12, respectively. The results show that even small ground-based robotic visual assistants can still provide views from the best manifolds for each affordance since all the manifolds whose value is in the 80th percentile of the manifold value range for the given affordance reach all the way to the ground except for manipulability (for which the three best manifolds reach the ground but the fourth-best does not).

D. Actionable Rules for Robotic Visual Assistants

The results can be extracted into actionable rules for robotic visual assistants suitable for a human operator to follow, as shown in Fig. 14. Those rules are extracted by partitioning the viewpoints on the hemisphere into five cardinal directions corresponding to the viewpoint groups from Fig. 7 and computing the view value for each of those cardinal directions as the mean value of the member viewpoints. The desired view direction for each affordance is extracted by taking the cardinal directions whose value is in the 80th percentile of the cardinal direction value range for the given affordance.

E. Sensitivity of Results to Different Weights

The manifolds for all affordances except manipulability are not very sensitive to w_m and w_d weights used in the viewpoint value [see (4)] and manifold value [see (11)] definitions. This is tested by starting with $w_m = 1$ and decreasing w_m by 0.1 in each step until $w_m = 0.5$ (the other weight is always $w_d = 1 - w_m$). It does not make sense to decrease the w_m weight below 0.5 as the mean should be the dominant indicator of the viewpoint or manifold value. For reachability, the shape and relative ranking of the manifolds remain unchanged. For passability, the shape and relative ranking of the manifolds remain unchanged until $w_m = 0.6$. For $w_m = 0.6$ and $w_m = 0.5$, M^{P}_5 has one less viewpoint, otherwise, the shape and relative ranking is unchanged. For traversability, the shape and relative ranking of the manifolds remain unchanged until $w_m = 0.7$. Then, the manifolds start changing, however, the two best manifolds, M^{T}_1 and M^{T}_2, remain

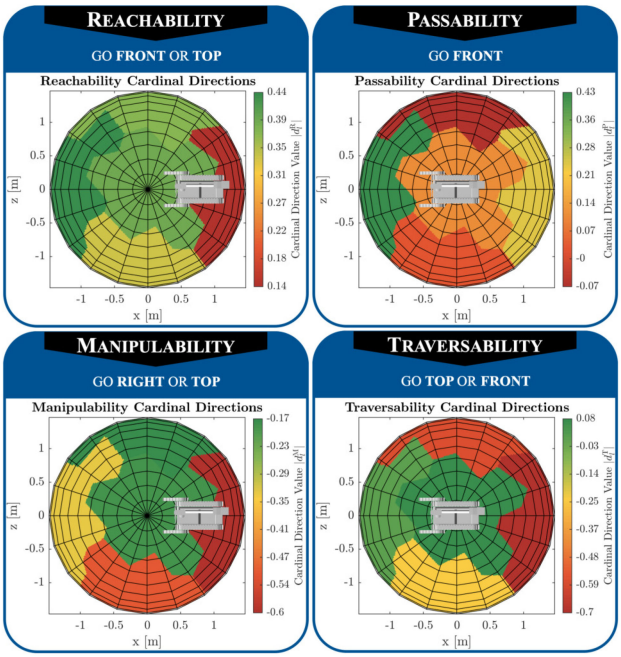

Fig. 14. Actionable rules for robotic visual assistants and the value for each of the five cardinal directions defined as $|d_l^a| = \text{mean}_{v_i \in d_l} |v_i^a|$, where d_l denotes lth cardinal direction.

unchanged until $w_m = 0.5$ and the best manifold, M_1^{T}, remains unchanged even for $w_m = 0.5$. Manipulability is the most sensitive to different weights, however, the first three manifolds, M_1^{M}, M_2^{M}, and M_3^{M}, remain unchanged in terms of their shape and relative order.

VIII. Summary

This work proposed a model of the value of different external viewpoints of a robot performing tasks. The model was developed using a psychomotor approach by quantifying the value of 30 external viewpoints for 4 affordances in a study with 31 expert robot operators using a computer-based simulator of two robots and clustering viewpoints of similar value into manifolds of viewpoints with equivalent value using agglomerative hierarchical clustering.

The results support the main postulation of the approach confirming the validity of the affordance-based approach by showing that there are manifolds of statistically significantly different viewpoint values, viewpoint values are statistically significantly dependent on the affordances, and viewpoint values are independent of the robot. The best manifold for each affordance provides a statistically significant improvement with a large Cohen's d effect size (1.1–2.3) in the performance and improvement in the performance variation over the worst manifold improving time by 14% to 59% and reducing errors by 87% to 100%.

This work creates the fundamental understanding of external viewpoints quality for four common affordances providing a foundation for ideal viewpoint selection; it is expected, but beyond the scope of this study that the application of the model to robotic visual assistants will likely reduce the cognitive workload on the primary operator. The model will enable autonomous selection of the best possible viewpoint and path planning for autonomous robotic visual assistants. One direction of future work would be to quantify view quality in a continuous matter rather than for a set of discrete viewpoints.

Acknowledgment

Preliminary results were published in [44] and [46]. The authors would like to thank Dr. S. C. Peres for providing valuable feedback and suggestions, and C. Oduola and M. Suhail for their significant work on the simulator and AWS infrastructure introduced in Section IV.

References

[1] R. R. Murphy, *Introduction to AI Robotics*. Cambridge, MA, USA: MIT Press, 2019.
[2] J. Dufek and R. Murphy, "Visual pose estimation of USV from UAV to assist drowning victims recovery," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, 2016, pp. 147–153.
[3] X. Xiao, J. Dufek, T. Woodbury, and R. Murphy, "UAV assisted USV visual navigation for marine mass casualty incident response," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 6105–6110.
[4] J. Dufek, X. Xiao, and R. Murphy, "Visual pose stabilization of tethered small unmanned aerial system to assist drowning victim recovery," in *Proc. IEEE Int. Symp. Saf., Secur. Rescue Robot.*, 2017, pp. 116–122.
[5] J. Dufek and R. Murphy, "Theoretical limitations of visual navigation of lifesaving USV using small UAS," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, 2018, pp. 1–7.
[6] J. Dufek and R. Murphy, "Visual pose estimation of rescue unmanned surface vehicle from unmanned aerial system," *Front. Robot. AI*, vol. 6, pp. 1–20, 2019. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2019.00042
[7] X. Xiao, J. Dufek, and R. Murphy, "Visual servoing for teleoperation using a tethered UAV," in *Proc. IEEE Int. Symp. Saf., Secur. Rescue Robot.*, 2017, pp. 147–152.
[8] X. Xiao, Y. Fan, J. Dufek, and R. Murphy, "Indoor UAV localization using a tether," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, 2018, pp. 1–6.
[9] X. Xiao, J. Dufek, and R. Murphy, "Benchmarking tether-based UAV motion primitives," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, 2019, pp. 51–55.
[10] X. Xiao, J. Dufek, M. Suhail, and R. Murphy, "Motion planning for a UAV with a straight or kinked tether," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 8486–8492.
[11] E. Guizzo, "Fukushima robot operator writes tell-all blog," 2011. Accessed: Apr. 24, 2020. [Online]. Available: https://spectrum.ieee.org/automaton/robotics/industrial-robots/fukushima-robot-operator-diaries
[12] R. R. Murphy, *Disaster Robotics*. Cambridge, MA, USA: MIT Press, 2014.
[13] S. Kawatsuma, M. Fukushima, and T. Okada, "Emergency response by robots to Fukushima-Daiichi accident: Summary and lessons learned," *Ind. Robot: Int. J.*, vol. 39, no. 5, pp. 428–435, 2012.
[14] G. T. McKee, B. G. Brooks, and P. S. Schenker, "Human-robot interaction for intelligent assisted viewing during teleoperation," in *Proc. 36th Annu. Hawaii Int. Conf. Syst. Sci.*, 2003, pp. 125–134.
[15] A. Roesler, "A new model for perspective: The role of point of observation in virtual and remote perspective-taking," Ph.D. dissertation, Dept. Ind. Syst. Eng., Ohio State Univ., Columbus, OH, USA, 2005. [Online]. Available: https://search.proquest.com/docview/305438884
[16] A. Morison *et al.*, "Integrating diverse feeds to extend human perception into distant scenes," in *Advanced Decision Architectures for the Warfighter: Foundation and Technology*. Columbus, OH, USA: Partners of the Army Research Laboratory Advanced Decision Architectures Collaborative Technology Alliance, 2009, pp. 177–200.
[17] A. M. Morison, "Perspective control: Technology to solve the multiple feeds problem in sensor systems," Ph.D. dissertation, Dept. Ind. Syst. Eng., Ohio State Univ., Columbus, OH, USA, 2010. [Online]. Available: https://search.proquest.com/docview/815248078

[18] A. M. Morison, T. Murphy, and D. D. Woods, "Seeing through multiple sensors into distant scenes: The essential power of viewpoint control," in *Human-Computer Interaction. Interaction Platforms and Techniques*. Toronto, ON, Canada: Springer, 2016, pp. 388–399.

[19] T. B. Murphy, "Apprehending remote affordances: Assessing human sensor systems and their ability to understand a distant environment," Ph.D. dissertation, Dept. Ind. Syst. Eng., Ohio State Univ., Columbus, OH, USA, 2013.

[20] T. B. Murphy and D. A. Morison, "Affordances as a means to assess human-sensor-robot performance," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, vol. 60, no. 1, pp. 1610–1611, 2016. [Online]. Available: https://doi.org/10.1177/1541931213601371

[21] T. Murphy and A. M. Morison, "Can I reach that? An affordance based metric of human-sensor-robot system effectiveness," in *Human-Computer Interaction. Theory, Design, Development and Practice*. Toronto, ON, Canada: Springer, 2016, pp. 360–371.

[22] T. B. Murphy, "Within reach: The contribution of dynamic viewpoint to the perception of remote environments," Ph.D. dissertation, Dept. Ind. Syst. Eng., Ohio State Univ., Columbus, OH, USA, 2017.

[23] A. M. Morison, D. D. Woods, and T. Murphy, *Human-Robot Interaction as Extending Human Perception to New Scales (Cambridge Handbooks Psychology)*. New York, NY, USA: Cambridge Univ. Press, 2015, pp. 848–868.

[24] D. D. Woods, J. Tittle, M. Feil, and A. Roesler, "Envisioning human-robot coordination in future operations," *IEEE Trans. Syst., Man, Cybern., Part C. (Appl. Rev.)*, vol. 34, no. 2, pp. 210–218, May 2004.

[25] S. Gatesichapakorn, M. Ruchanurucks, P. Bunnun, and T. Isshiki, "ROS-based mobile robot pose planning for a good view of an onboard camera using costmap," in *Proc. 10th Int. Conf. Inf. Commun. Technol. Embedded Syst.*, 2019, pp. 1–6.

[26] G. Yang, S. Wang, J. Yang, and B. Shen, "Viewpoint selection strategy for a life support robot," in *Proc. IEEE Int. Conf. Intell. Saf. Robot.*, 2018, pp. 82–87.

[27] X. Bonaventura, M. Feixas, M. Sbert, L. Chuang, and C. Wallraven, "A survey of viewpoint selection methods for polygonal models," *Entropy*, vol. 20, no. 5, pp. 1–22, 2018. [Online]. Available: https://www.mdpi.com/1099-4300/20/5/370

[28] P. Kurtser and Y. Edan, "The use of dynamic sensing strategies to improve detection for a pepper harvesting robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 8286–8293.

[29] R. Sato, M. Kamezaki, S. Niuchi, S. Sugano, and H. Iwata, "Derivation of an optimum and allowable range of pan and tilt angles in external sideway views for grasping and placing tasks in unmanned construction based on human object recognition," in *Proc. IEEE/SICE Int. Symp. Syst. Integration*, 2019, pp. 776–781.

[30] E. Dima *et al.*, "View position impact on QoE in an immersive telepresence system for remote operation," in *Proc. 11th Int. Conf. Qual. Multimedia Experience*, 2019, pp. 1–3.

[31] R. Sato, M. Kamezaki, M. Yamada, T. Hashimoto, S. Sugano, and H. Iwata, "Experimental investigation of optimum and allowable range of side views for teleoperated digging and release works by using actual construction machinery," in *Proc. IEEE/SICE Int. Symp. Syst. Integration*, 2019, pp. 788–793.

[32] J. d. León, M. Garzón, D. Garzón, E. Narváez, J. D. Cerro, and A. Barrientos, "From video games multiple cameras to multi-robot teleoperation in disaster scenarios," in *Proc. Int. Conf. Auton. Robot Syst. Competitions*, 2016, pp. 323–328.

[33] S. Kiribayashi, K. Yakushigawa, and K. Nagatani, "Design and development of tether-powered multirotor micro unmanned aerial vehicle system for remote-controlled construction machine," in *Field and Service Robotics*. Zurich, Switzerland: Springer, 2018, pp. 637–648.

[34] D. Nicolis, M. Palumbo, A. M. Zanchettin, and P. Rocco, "Occlusion-free visual servoing for the shared autonomy teleoperation of dual-arm robots," *IEEE Robot. Automat. Lett.*, vol. 3, no. 2, pp. 796–803, Apr. 2018.

[35] A. Gawel, Y. Lin, T. Koutros, R. Siegwart, and C. Cadena, "Aerial-ground collaborative sensing: Third-person view for teleoperation," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, 2018, pp. 1–7.

[36] S. Chikushi *et al.*, "Automated image presentation for backhoe embankment construction in unmanned construction site," in *Proc. IEEE/SICE Int. Symp. Syst. Integration*, 2020, pp. 22–27.

[37] S. Samejima, K. Fozilov, and K. Sekiyama, "Visual support system for remote control by adaptive ROI selection of monitoring robot," *ROBOMECH J.*, vol. 5, no. 1, pp. 1–21, Mar. 2018. [Online]. Available: https://doi.org/10.1186/s40648-018-0103-0

[38] J. Thomason *et al.*, "A comparison of adaptive view techniques for exploratory 3D drone teleoperation," *ACM Trans. Interactive Intell. Syst.*, vol. 9, no. 2/3, pp. 17: 1–17:19, Mar. 2019. [Online]. Available: http://doi.acm.org/10.1145/3232232

[39] D. Rakita, B. Mutlu, and M. Gleicher, "Remote telemanipulation with adapting viewpoints in visually complex environments," in *Proc. Robot.: Sci. Syst. XV*, pp. 1–10, Jun. 2019. [Online]. Available: http://par.nsf.gov/biblio/10104548

[40] J. J. Gibson, *The Ecological Approach to Visual Perception: Classic Edition*. New York, NY, USA: Psychology Press, 2014.

[41] E. W. Hallford, "Sizing up the world: The body as referent in a size-judgment task," Ph.D. dissertation, Dept. Psychol., Ohio State Univ., Columbus, OH, USA, 1984.

[42] X. Xiao, J. Dufek, and R. Murphy, "Explicit motion risk representation," in *Proc. IEEE Int. Symp. Safety, Security, Rescue Robot.*, 2019, pp. 278–283.

[43] X. Xiao, J. Dufek, and R. R. Murphy, "Robot motion risk reasoning framework," 2019, *arXiv:1909.02531*. [Online]. Available: http://arxiv.org/abs/1909.02531

[44] X. Xiao, J. Dufek, and R. R. Murphy, "Autonomous visual assistance for robot operations using a tethered UAV," in *Field and Service Robotics*, G. Ishigami and K. Yoshida, Eds. Singapore: Springer, 2021, pp. 15–29.

[45] X. Xiao, J. Dufek, and R. Murphy, "Explicit-risk-aware path planning with reward maximization," 2019, *arXiv:1903.03187*. [Online]. Available: http://arxiv.org/abs/1903.03187

[46] X. Xiao, J. Dufek, and R. R. Murphy, "Tethered aerial visual assistance," 2020, *arXiv:2001.06347*. [Online]. Available: https://arxiv.org/abs/2001.06347

[47] X. Xiao, J. Dufek, and R. R. Murphy, "Robot risk-awareness by formal risk reasoning and planning," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2856–2863, Apr. 2020.

[48] C. D. Michener and R. R. Sokal, "A quantitative approach to a problem in classification," *Evolution*, vol. 11, no. 2, pp. 130–162, 1957. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1558-5646.1957.tb02884.x

[49] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.*, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/03610927408827101