Human Perception-Optimized Planning
for Comfortable VR-Based Telepresence

Israel Becerra[1,2], Markku Suomalainen[3], Eliezer Lozano[1], Katherine J. Mimnaugh[3],
Rafael Murrieta-Cid[1] and Steven M. LaValle[3]

Abstract— This paper introduces an emerging motion planning problem by considering a human that is immersed into the viewing perspective of a remote robot. The challenge is to make the experience both effective (such as delivering a sense of presence) and comfortable (such as avoiding adverse sickness symptoms, including nausea). We refer this challenging new area as *human perception-optimized planning* and propose a general multiobjective optimization framework that can be instantiated in many envisioned scenarios. We then consider a specific VR telepresence task as a case of human perception-optimized planning, in which we simulate a robot that sends 360 video to a remote user to be viewed through a head-mounted display. In this particular task, we plan trajectories that minimize VR sickness (and thereby maximize comfort). An A* type method is used to create a Pareto-optimal collection of piecewise linear trajectories while taking into account criteria that improve comfort. We conducted a study with human subjects touring a virtual museum, in which paths computed by our algorithm are compared against a reference RRT-based trajectory. Generally, users suffered less from VR sickness and preferred the paths created by the presented algorithm.

I. INTRODUCTION

In the last few years, the arrival of consumer Virtual Reality (VR) products has enhanced the level of immersion that most people can experience through a robotic platform. This is an unprecedented opportunity to make people feel present in a remote or artificial environment along with the actuation provided by robotic platforms (see Fig. 1). This allows people to interact with each other over distances as more than a face on a screen, in so-called *mobile robotic telepresence*, which has been shown to be a superior means of communication over simple videoconferencing [18], [25]; possible use cases include attending conferences or business meetings, and elderly care [1]. This highly immersive mode of human-robot interaction brings challenging new motion planning aspects.

We first present a mathematical framework for *human perception-optimized planning*, in which unprecedented level

[1]I. Becerra, E. Lozano and R. Murrieta-Cid are with Centro de Investigación en Matemáticas (CIMAT), 36023 Guanajuato, México, {israelb, eliezer.lozano, murrieta}@cimat.mx
[2]I. Becerra is with Consejo Nacional de Ciencia y Tecnología, CONA-CyT, 03940 Mexico City, México
[3]M. Suomalainen, K. J. Mimnaugh and S. M. LaValle are with Center of Ubiquitous Computing, Faculty of Information Technology and Electrical Engineering, University of Oulu, Finland. {Markku.Suomalainen, Katherine.Mimnaugh, Steven.Lavalle}@oulu.fi

Fig. 1. An illustration of telepresence in which the robot is equipped with a 360 camera and the user with an HMD. The left picture was taken during our user study.

of human factors must be considered in the motion planning problem. In this framework one of the most challenging problems is to guarantee user comfort as several of the user's senses are stimulated with artificial or remote experiences, while also taking into account classic motion planning criteria and other people in the presence of the robot. We propose a general and formal definition of such framework, formulated as a Multi-Objective Optimization (MOO) problem, which can be instantiated in a variety of tasks.

We then present a concrete VR telepresence task as an instance of human perception-optimized planning, in which we simulate a robot that streams 360 video to a user's head-mounted display (HMD). Although the use of an HMD may bring on adverse side effects, it has been shown superior in some contexts such as search-and-rescue operations requiring detailed object identification and depth perception [20], collaborative assembly tasks [12], and increased stability while teleoperating a wheelchair [7]. This makes HMD-based teleoperation an appealing research direction. Even though there are known methods for reducing VR sickness, their downside is that they often reduce the feeling of *presence* [35], which is the feeling of actually being at the location seen on the screen [30]. Increasing the feeling of presence has the potential to resolve some issues of telepresence, such as perceived difficulty of communication [32]. Because the susceptibility to VR sickness varies greatly within a population [27], applying the sickness reduction techniques to all users should not be considered. Thus, in this work we consider finding the *Pareto front* of the MOO problem, which means the set of all solutions for which there are none

better in terms of all criteria.

This paper has two main parts. The first one is to define the general framework for human perception-optimized planning as a MOO problem with a Pareto front, which allows multiple future research directions and usage with any motion planning method. This is done by defining four classes of criteria to help researchers make sure they have taken all of the required parts of a problem into account. The second part is to use the framework to define a VR telepresence task as a concrete instance of human perception-optimized planning. For this particular instance, we propose computing piecewise-linear trajectories, which are motivated by falling in line with the *sensory conflict theory* [26] of motion sickness.

Finally, in preliminary experiments, we show with trials on human subjects that the Pareto trajectories designed by the presented algorithm are preferred over trajectories created by a typical motion planning algorithm, the RRT [16] (even when selecting an optimized trajectory). In such trials, the remote video streaming performed by the robot is simulated, allowing us to have control over the experimentation. The actual tests with a physical robot are left for future work, where issues such as communication delays, feedback control, and robot localization must be considered. Nonetheless, the simulation study in this paper is an important step toward validation of the planning methods.

Previous work

MOO issues arise often in robotics because of common tradeoffs between safety and task efficiency. However, this is typically solved by simply choosing the weights, or priorities, beforehand. In contrast, *Pareto optimization* finds all possible solutions, often called the Pareto front, in which one criterion cannot be improved without degrading another one. Many Pareto-optimization works appear in robotics [15], [37], [38]. Pareto optimization is a natural choice for human perception-optimized planning because the prioritizing or weighting of the criteria can vary heavily due to the high variance of susceptibility to cybersickness in population.

In this paper, we consider a VR telepresence task as a particular instance of human perception-optimized planning. In the area of robotics, the human telepresence task has been addressed in works such as [17], [24], with the goal to achieve the remote presence of a human in a physical space. Whereas most works in telepresence consider only a 2-D screen, there are few works who consider using an HMD as well. Oh et al. [23] considered using a telepresence robot streaming video from 360 camera to an HMD on a facility tour. Heshmat et al. [8] compared telepresence between a 2-D screen and an HMD, and found that people prefer the ability to look around with the HMD. Zhang et al. [39] researched the use of redirected walking when using an HMD for teleoperation, and Stotko et al. [33] built a model from the environment to be observed by the user while the robot moved. However, in [33], [39] the issue of VR sickness was avoided by using less scalable user interfaces, and in [8], [23] it was not considered at all.

The term VR sickness, often also called cybersickness, is used to refer to Visually Induced Motion Sickness (VIMS) in the context of Virtual Reality [5], [22]. To be more specific, VIMS is a particular type of Motion sickness (MS) that may occur without the person physically moving but while they are observing motion. VIMS can appear under visual stimulation present in movie theatres [28], virtual simulators [6] and video games. It can result in symptoms such as cold sweat, dizziness, headaches, nausea, and even vomiting. Since MS and VIMS share many common characteristics, classical MS theories that incorporate visual components can be used to try to explain and address VIMS, and therefore, VR sickness. One theory is that the origin of MS is a negative reinforcement system to avoid postural instability [2]. Other works as [29] explain MS as a function of the vestibular detection of stimuli that would be disruptive to digestion. Some theories even suggest that MS serves as a mechanism to avoid poisoning [34]. Nevertheless, the most accepted and cited MS theory is the *sensory conflict theory* [4], [26]. This theory attributes MS to the mismatch between optical flow perceived by the eyes, the vestibular system, and/or the somatosensory senses (non-vestibular proprioceptive senses of skin, muscles, and joints). This last theory is the one we adopt in the presented work.

Consider then our scenario, in which a stationary human user is wearing an HMD, and a remote robot is used to stream 360 video to the user through the HMD. The main potential discomfort in such a system comes from the symptoms of VR sickness. More precisely, assume that the user is seated wearing the HMD; when the robot moves and transmits views from changing viewpoints, the vestibular system reports that the user is motionless, but the user's vision system reports to his brain that they are moving, which might yield vection. *Vection* is the illusion of self motion when no movement is taking place, and is believed to be an important cause for VR sickness, as vection involves an intrinsic sensorial conflict that might result in symptoms such as dizziness, nausea, and even vomiting. Several factors affect vection sensitivity [14]. Some examples include the distance from the center view, spatial frequency of the displayed images, prior knowledge (knowing beforehand what kind of motion should be perceived), and exposure time to the optical flow.

II. HUMAN PERCEPTION-OPTIMIZED PLANNING

We define as *human perception-optimized planning* the generation of a collision-free trajectory for a sensing-system that generates a perceptual stimulus to an interfaced user, while ensuring user's comfort. If the sensing system is attached to a mobile robot platform, both the sensor and the platform may have separate Degree of Freedoms (DoFs); consider, for example, a camera attached to a pan-tilt-unit, or even a robot arm, on top of a wheeled platform. This decoupling allows assigning different requirements for each set of DoFs; it has been shown that decoupling the viewing angle and motion of the vehicle improves teleoperation [9].

Human perception-optimized planning can be considered as an upper layer to a motion planning task; the perceptual

stimuli is planned to optimize a set of criteria. Inherited by the motion planning aspect of the task, the path itself can be required to optimize certain aspects of the movement, for example, the travelled distance. However, these criteria may be contradicting and depend on personal preferences, and thus care must be taken on how to prioritize the criteria, naturally leading to the formulation of Pareto optimization.

To formalize the *human perception-optimized planning* problem, we proceed to introduce some notation and basic concepts. The physical system consists of a mobile robot base B and a (possibly actuated) sensor E attached to the base, with both moving in the Euclidean workspace W. Let $O \subset W$ be the obstacle region, C_E the configuration space of E and C_B the configuration space of B, thus making the configuration space of the whole system $C = C_E \times C_B$. Finally, C_{obst} is the set of configurations in which the interior of the system geometric model, placed at configuration $q \in C_{obst}$, intersects O.

Let X denote the state space, which is formed as the Cartesian product of C and a compact space that covers time derivatives of configuration. Let $X_{obst} = \{\mathbf{x} \in X \mid q \in C_{obst}\}$, and $X_{valid} = X \setminus X_{obst}$. Furthermore, let $\sigma \in \Sigma_{valid}$ be a continuous, collision-free trajectory, with $\sigma : [0, T] \to X_{valid}$, in which $\sigma(0)$ is the initial state $\mathbf{x}_{init}$, and $\sigma(T)$ is the goal state $\mathbf{x}_{goal}$. Let Σ_{valid} denote the set all such trajectories (assuming $\mathbf{x}_{init}$ and $\mathbf{x}_{goal}$ are fixed). In some cases, Σ_{valid} may be further constrained to include only trajectories that satisfy a control model of the form $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u})$, with input $\mathbf{u}$ drawn from a compact set U.

A trajectory that solves the *human perception-optimized planning* will be required to optimize certain criteria. For a better classification of the problem, we define four classes, J_P, J_C, J_R and J_O, which group the criteria belonging to different aspects of the problem. This classification can be used to, for example, make sure that no aspect of the problem is ignored or overemphasized. Each individual criterion is defined as a cost functional $J_i : \Sigma_{valid} \to \mathbb{R}_{>0}$. First, the class J_P includes criteria defining the *performance*, which depends on the application and refers to keeping the intended functionality of the system. For example, if the robot is equipped with a manipulator, this could correspond to keeping the orientation such that the manipulator can be used safely (e.g., not have it face a wall). In a telepresence scenario, this could correspond to the user retaining spatial orientation and the sense of presence. J_C measures the comfort of the interfaced user while exposed to the stimulus obtained from the system, while moving via σ; in the case of an HMD, this would correspond to criteria mitigating VR sickness. J_R is a function that considers the *robot* motion; this includes mainly traditional motion planning criteria, such as path length, distance to objects, power consumption, etc. Lastly, J_O considers *others*, for instance, other humans or moving bodies in the vicinity. These criteria can be related to the human-aware motion planning [13], in which behaviors such as not walking between two conversing people are considered. Thus, the task is to compute some $\sigma \in \Sigma_{valid}$ that simultaneously minimizes the multiple criteria given by

the vector of costs:
$$
\begin{aligned}
J(\sigma) &= (J_P(\sigma), J_C(\sigma), J_R(\sigma), J_O(\sigma)) \\
&= (J_1(\sigma), J_2(\sigma), \ldots, J_k(\sigma)),
\end{aligned}
$$
in which $J_1 \ldots J_k$ cover all individual criteria from J_P, J_C, J_R and J_O. The classes can be thought of as a manner to organize the costs, hence, the classes are vectors themselves. We note that this classification of criteria is meant to clarify the definition of a problem, and make sure that researchers and engineers who design human-based motions consider all aspects of the problem: criteria such as the distance to objects can be considered both as safety (J_R) or performance (J_P); with a 360 camera, the performance deteriorates heavily if the camera is too close to an object.

Usually in a MOO problem a single solution is found by weighting the criteria according to their importance. However, when individual human physiology can have a strong impact on a desired weighting, scalarization of the problem too early should be avoided. This is especially important for the telepresence application because it has been shown that VR sickness and presence have a negative correlation [35], and thus degrading the presence for people who do not suffer from VR sickness severely deteriorates their situational awareness and communication ability. A natural solution is to find the Pareto front [3], in other words, solutions that cannot be improved in any of the objectives without degrading at least one of the other objectives. Mathematically, this is defined through the concept of (Pareto) dominance: a trajectory σ_1 *dominates* trajectory σ_2, denoted as $\sigma_1 \preceq \sigma_2$, if $J_i(\sigma_1) \leq J_i(\sigma_2) \ \forall i \in \{1, 2, \ldots, k\}$, and $\exists j \in \{1, 2, \ldots, k\}$ such that $J_j(\sigma_1) < J_j(\sigma_2)$. Finally, based on the previous concepts, the general problem formulation is given next.

Human perception-optimized planning: *Given the general motion planning formulation above, and a cost functional* $J(\sigma) = (J_P(\sigma), J_C(\sigma), J_R(\sigma), J_O(\sigma))$, *find a set of all trajectories (up to cost-vector equivalence):*
$$
\{\sigma^* \in \Sigma_{valid} \mid \nexists \sigma \in \Sigma_{valid} \text{ for which } \sigma \preceq \sigma^*\}. \quad (1)
$$

Note that this formulation does not consider a single trajectory to be a solution. This could be accomplished by simply formulating and optimizing a scalar, linear combination of all of the objectives; however, we want to present the set of Pareto-optimal solutions so that the system, together with users, could select particular trajectories dynamically during execution. It is important to offer this because to the high variability of human subject sensitivities and environmental conditions that arise during execution.

The next sections apply this formulation to an illustrative VR telepresence task.

III. CASE STUDY: VR TELEPRESENCE

The concept of telepresence is attributed to Marvin Minsky, pioneer of artificial intelligence [21]. In the present work, we refer to VR telepresence as the set of technologies that allow human users interfaced with VR equipment to feel as they were present in a remote location, and even allow them to interact in that location through the use of

teleoperated robots. Particularly, we will consider that there is a robot in a remote location equipped with a 360 camera, and that it is streaming video to an HMD worn by a human user at a different location (see Fig. 1). It is assumed that the human user has control over the robot's goal (essentially, any location in X_{valid}), but that a motion planner computes the trajectories to reach the goal; therefore, solving a human perception-optimized planning problem.

In the next subsections we present the specifics of the criteria used in this case study for human perception-optimized planning, and the actual planner that will be used to compute the required trajectories.

A. System model

The considered system consists of a robotic base moving on a plane with a 360 camera mounted on top of it, streaming video to an HMD worn by a user. The robotic base is a differential drive robot (DDR). The camera will be fixed on top of the DDR, hence, the configuration space of the whole system will be the one of the robotic base, namely, $C = C_B$. Considering extra degrees of freedom (C_E) for an actuated or filtered omnidirectional camera is left for future work.

The state of the system is given by the tuple (x, y, θ, v_r, v_l). Let x and y be the coordinates of the reference point located between the DDR wheels, and θ the robot's heading. Variables v_r and v_l denote the velocities of the contact point of the right and left wheels with the floor, respectively. The system is controlled through the right and left wheels' translational accelerations, a_r and a_l, with $a_r, a_l \in [-a_{max}, a_{max}]$, $a_{max} > 0$. Considering b as the distance between the robot's reference point and the wheels, we obtain the next state transition equation that models our VR telepresence system:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{v}_r \\ \dot{v}_l \end{pmatrix} = \begin{pmatrix} \frac{v_r + v_l}{2} \cos\theta \\ \frac{v_r + v_l}{2} \sin\theta \\ \frac{v_r - v_l}{2b} \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} a_r + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} a_l. \quad (2)$$

B. Modeling of user comfort cost functional

A critical aspect on the VR telepresence task is the user's comfort, which is mainly affected by the experienced VIMS. Nonetheless, other performance issues need to be addressed; with a 360 camera, the performance deteriorates heavily if the camera is too close to an object. In the present work, to compute solution trajectories for the human-perception-planning problem in the context of VR telepresence, we will mainly focus on the performance and user's comfort aspects of the problem; therefore, the cost functional will be of the form $J(\sigma) = (J_P(\sigma), J_C(\sigma))$. This allows a controlled experiment that concentrates on the sickness and preference; new criteria must be thoroughly controlled and researched, before they are accepted as a part of a complete human perception-optimized planning.

Regarding the performance costs J_p, due to the 360 camera requirements, it is desirable to keep a ball of radius r

around the 360 camera unobstructed. Consequently, we define a function $\varphi(\mathbf{p})$ that measures the obstructed percentage of such a ball centered at the camera position $\mathbf{p}$. Eq. (3) defines $\varphi(\mathbf{p})$, in which for a measurable set $A \subset \mathbb{R}^3$, $\mu(A)$ is the volume of A, $S(\mathbf{p}, r)$ is a ball centered at $\mathbf{p}$ with radius r, and O is the obstacle region in W:

$$\varphi(\mathbf{p}) = \frac{\mu(S(\mathbf{p}, r) \cap O)}{\mu(S(\mathbf{p}, r))}. \quad (3)$$

Using function $\varphi(\mathbf{p})$, we define the function $V(\sigma)$ as in Eq. (4), which is an average over all trajectory $\sigma : [0, T] \to X_{valid}$ of the percentage of obstructed volume of a ball of radius r around the 360 camera:

$$\mathcal{V}(\sigma) = \frac{1}{T} \int_0^T \varphi(\mathbf{p}) dt. \quad (4)$$

This functional is aimed at preferring trajectories in which the ball around the 360 camera is not cluttered, allowing proper functioning of the camera. Concerning user's comfort J_C, from a sensory conflict theory perspective, the experienced VIMS comes from the conflict between the user's vision and vestibular system. The vestibular system is composed of two main organs, the otoliths and the semicircular canals. The otoliths sense linear acceleration and the semicircular canals angular acceleration. Under that premise, presenting the visual stimuli corresponding to following a curved path would evoke potential sensory conflict with the otoliths, due to the presence of linear accelerations (for instance, the components of centripetal acceleration). Moreover, visual stimuli resulting from rotational movement can also evoke sensory conflict with the semicircular canals. Under that rationale, in the present work, we propose to move along *piecewise linear* paths, in addition to reducing the number of line segments comprising them. The DDR will be required to apply straight line motions with its heading pointing tangentially to the line segments, and apply rotations in place at line segment transitions to redirect its heading with regard to the next segment. Note that following line segments, the total time of conflict with the otoliths can be reduced; as shown in [36], performing a fixed amount of rotation with greater speed can be beneficial in preventing VR sickness, as it reduces the total time of conflict. Additionally, reducing the number of segments reduces the number of poses at which conflict with the semicircular canals takes place. Indeed, the number of transitions between line segments, $N(\sigma)$, is set to be part of our cost functional associated to user comfort. Even more, there is evidence that rotational motion is the most evocative of VR sickness [10].

Additionally, it is also desirable to minimize the length of the path (see Eq. (5)) because there is a direct relation between the path length and the time of exposure to potential sensory conflict due to motion. Thus, the user's comfort class is set as $J_C(\sigma) = (N(\sigma), D(\sigma))$, and

$$D(\sigma) = \int_0^T \sqrt{\dot{x}^2(t) + \dot{y}^2(t)} dt. \quad (5)$$

The resulting cost vector is defined as

$$J(\sigma) = (\mathcal{V}(\sigma), N(\sigma), D(\sigma)). \quad (6)$$

C. Motion planner

We present a planner that addresses the human perception-optimized planning problem formulated in Section II. Although that covers general trajectories, there are several advantages for the telepresence setup in restricting the search space to piecewise linear paths: reduction of VR sickness, planning simplicity, and potential for improved retaining of spatial orientation. The piecewise-linear path requirement makes it suitable to consider a regular grid representation of the configuration space $C = SE(2)$. Considering a grid will naturally result in paths composed of linear segments.

The regular grid is modeled as a directed graph, $G = (V, E)$, in which each node, $v_{\mathbf{p},\theta} \in V$, is labeled with a position $\mathbf{p} = (x, y)$ on the plane, and a given orientation θ. The positions $\mathbf{p}$ are equally spaced throughout the x and y coordinates of the plane using a step δ, and the orientations θ lie in the set $\{0°, 45°, ..., 270°, 315°\}$, to preserve eight-neighbor connectivity in the plane (see Fig. 2). Transition between elements in the grid are defined through the following edge definition. First, let $e(v_{\mathbf{p},\theta} v_{\mathbf{p}',\theta'})$ denote an edge from node $v_{\mathbf{p},\theta}$ toward $v_{\mathbf{p}',\theta'}$–the arguments of $e(v_{\mathbf{p},\theta} v_{\mathbf{p}',\theta'})$ will be dropped when convenient. Such edge will exist if $\mathbf{p}' = \mathbf{p}$ and $\theta' \neq \theta$ (refer to them as **Type-A** edges); or if $\theta' = \theta$ and $v_{\mathbf{p}',\theta'}$ is a neighbor of $v_{\mathbf{p},\theta}$ under an 8-connectivity in the plane (refer to them as **Type-B** edges). See Fig. 2 for examples of **Type-A** and **Type-B** edges.

Each edge, $e(v_{\mathbf{p},\theta}, v_{\mathbf{p}',\theta'})$, will have associated to it a non-negative cost vector, $w(e) = (w_1(e), w_2(e), w_3(e)) \in \mathbb{R}^3$. Its first element, $w_1(e)$, is associated to the cost $\mathcal{V}$ that evaluates the obstructions around the 360 camera, and it is simply set as $w_1(e) = \varphi(\mathbf{p}')$. The cost $w_2(e)$ is associated to N, the number of turns performed by the DDR. It is set as $w_2(e) = 1$ for **Type-A** edges, and as $w_2(e) = 0$ for **Type-B** edges. The last element, $w_3(e)$, is associated to D, the traveled distance. For **Type-A** edges $w_3(e) = 0$. For **Type-B** edges, $w_3(e) = \delta$ if $\theta \in \{0°, 90°, 180°, 270°\}$, and $w_3(e) = \sqrt{2}\delta$ if $\theta \in \{45°, 135°, 235°, 315°\}$. See Fig. 2 for cost vector examples.

To compute the actual Pareto optimal trajectories, σ^*, we use the multiobjective variant of the A* algorithm presented in [19]. That variant is an extension of A* to the MOO case, preserving path selection and expansion as the basic operations, as opposed to MOA* [31] that preserves node selection and expansion instead. The path selection results into substantial savings in memory over MOA*. For a given directed graph and set of edges' costs, the algorithm from [19] computes the set of all non-dominated solutions whenever this set is finite and nonempty. Fig. 3 shows some sample trajectories that we were able to compute with the aforementioned algorithm for our problem modeling. Lastly, based on the findings from [36] to reduce VR sicknes, the DDR is required to follow the path with constant speed; hence, accelerations only take place at the beginning and at the end of the straight line motions and rotations in place.
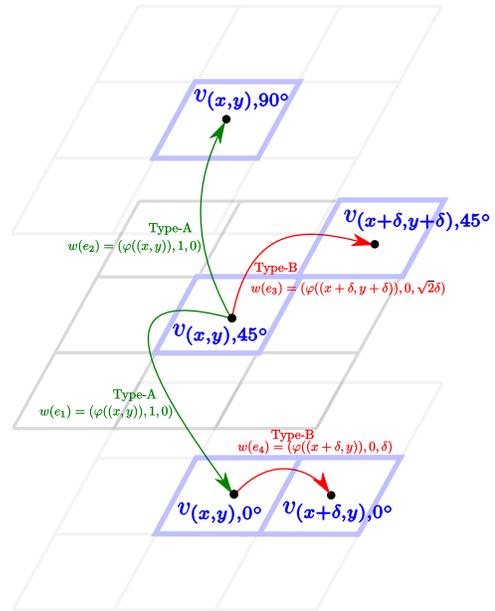

Fig. 2. Examples of the connectivity and edge weighting of the graph $G = (V, E)$ modeling C. Examples of Type-A and Type-B edges are shown, along with their respective cost vectors. Type-A edges correspond to the DDR applying rotations in place, and Type-B edges correspond to the DDR applying straight line motions.

IV. EXPERIMENTS

A. Experimental setup

We performed a user study in a laboratory using a completely virtual museum environment built with Unity 3D (see Fig. 4). We first ran the planner from Section III-C on a 2D projection of the museum environment. From the set of Pareto-optimal trajectories, we hand-picked two: one that minimizes the number of rotations, and another that reduces the distance traveled. The RRT trajectory was selected from 1000 solutions produced by RRT reruns, preferring the one with the least number of curvature sign changes. The three chosen trajectories are shown in Fig. 4. A 360 video recording of a mobile robot traversing each of these trajectories in the environment was created in Unity, so that the subjects could rotate their heads and look around in the environment (as they would in a real setting), but could not control the robot's trajectory. To compare how subjects felt about the trajectories, we performed a within-subject user study with 36 participants. The study took place in a research lab at the University of Oulu (Fig. 1, left picture). The subjects were equally balanced by gender with 18 females and 18 males. Their ages ranged from 20 to 44 with a mean age of 28.25 years. The presentation of each of the three videos was fully counterbalanced to counteract any potential ordering effects; therefore, three females and three males each saw one of the six combinations of video presentation orders.

Subjects were first asked to sign a consent form and fill in the baseline Simulator Sickness Questionnaire (SSQ) [11], after which instructions were given and the first video was played. Subjects were instructed not to pay attention to the pieces of art they saw on the path (as there were two different
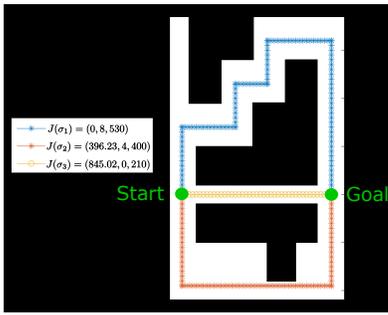
Fig. 3. Three Pareto optimal solutions σ_i for the VR telepresence problem are shown. The associated costs $J(\sigma_i) = (\mathcal{V}(\sigma_i), N(\sigma_i), D(\sigma_i))$ are also displayed. The solutions were computed considering $r = 2$ for the 360 camera. In the stat and goal configurations, the robot's heading is aligned with the positive direction of the x-axis. Path σ_1 passes through the upper part of the environment maintaining the ball around the 360 camera completely unobstructed, but at the cost of generating a long path requiring 8 rotations in place. Path σ_3 passes through the narrow passage in the middle greatly deteriorating clearance around the 360 camera, but generates the shortest path to reach the goal while not needing rotations. Path σ_2 is in a middle ground in terms of cost.
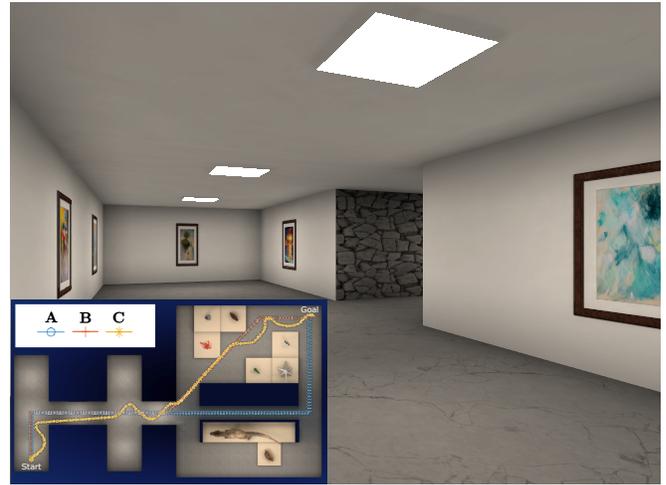

Fig. 4. A screenshot from the museum environment used in the user study. A top-down view of the museum is also shown along with three tested trajectories. Pareto trajectory that minimizes the number of rotations is labeled as **A**. Pareto trajectory that focus on reducing distance is labeled as **B**. The tested RRT trajectory is labeled as **C**.

homotopy classes, as shown on Fig. 4). After each of the three videos, the subjects were asked to fill in an SSQ and another questionnaire with 6-point Likert scale questions regarding their comfort, retention of sense of orientation, and perception of closeness to walls and objects. Finally, after the last video, subjects were asked to select which of the three videos they preferred and which was the most comfortable. Each video lasted between 1min and 1min 30s.

The visual features in the Pareto least turns path were different in the second half of the video from the features shown in the second half of the Pareto shortest path and the RRT (see Fig. 4 top-view mini-map). The Pareto least turns path passed through a hallway with blank walls, while the other paths passed through a room filled with sculptures. Despite this difference, the three paths were constructed to take the user from the same initial state $\mathbf{x}_{init}$ to the same goal state $\mathbf{x}_{goal}$.

B. Results

All tests were run with a 95% confidence interval and two-tailed significance levels set to 0.05. Shapiro-Wilk tests found a departure from normality for the response distributions, so non-parametric tests were used. Significance values were adjusted with Bonferroni correction for multiple tests.

A Wilcoxon Signed-Ranks test indicated that the Pareto least turns path did not result in a statistically significant increase in SSQ scores from the baseline ($Mdn = 5.61$) to the post-test ($Mdn = 9.35$), Z = -0.501, $p = .616$. The Pareto shortest path resulted in a nearly statistically significant increase in SSQ scores from the baseline to the post-test ($Mdn = 14.96$), Z = -1.931, $p = .054$. However, the RRT path resulted in a statistically highly significant increase in SSQ scores from the baseline to the post-test ($Mdn = 20.57$), Z = -3.328, $p = .001$.

A Kruskal-Wallis H test was performed to compare the post-treatment mean SSQ scores for each of the paths.

A statistically significant difference was found, $\chi^2(2) = 9.035$, $p = .011$. Post-hoc pairwise comparisons revealed a statistically significant difference between the Pareto least turns path and the RRT ($p = .008$). The Pareto least turns and Pareto shortest path were not significantly different, ($p = .566$), and the Pareto shortest path was not significantly different from the RRT ($p = .277$).

Figs. 5 and 6 show the distribution of responses regarding users' comfort and preference based on answers obtained from the questionnaire asking them to select one of the three paths. Fig. 7 presents the total weighted SSQ scores recorded after watching the video of each of the paths. Only the Pareto least turns and the RRT paths had a statistically significant difference.

C. Discussion

The results of this study show that the most comfortable path is the Pareto least turns path. Viewing the video of this path resulted in only a small increase in the SSQ total weighted symptom scores from the baseline pre-test to the post-test, which was not a statistically significant difference. The Pareto least turns path was also significantly more comfortable than the RRT path. This supports our hypothesis that the number of turns plays an important role in the users' comfort when viewing these paths in a virtual reality headset. This finding is consistent with the results from a previous study [10], where it was found that rotational movement is the most evocative of VR sickness.

Concerning the Pareto shortest path, our results found that this path was more comfortable than the RRT path but less comfortable than the Pareto least turns path. The median SSQ total weighted score at the baseline pre-test was 5.61 and the median scores after watching each video (see Fig. 7) were 9.35 for the Pareto least turns, 14.96 for the Pareto shortest path and 20.57 for the RRT. From the Wilcoxon test
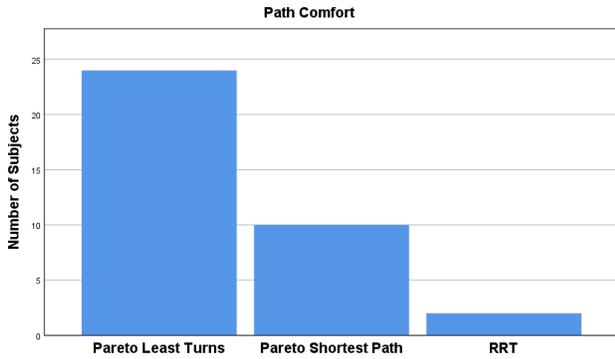
Fig. 5. Responses to the question: Of the three paths, which one was the most comfortable?
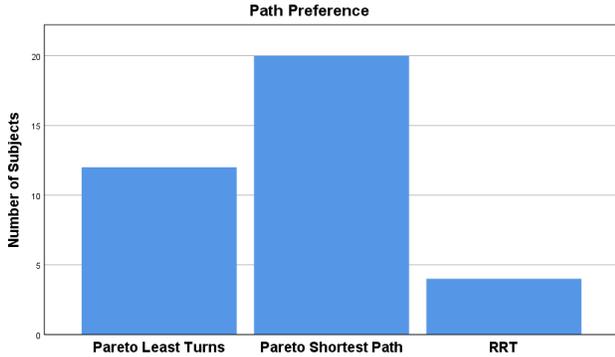

Fig. 6. Responses to the question: Of the three paths, which one did you prefer?
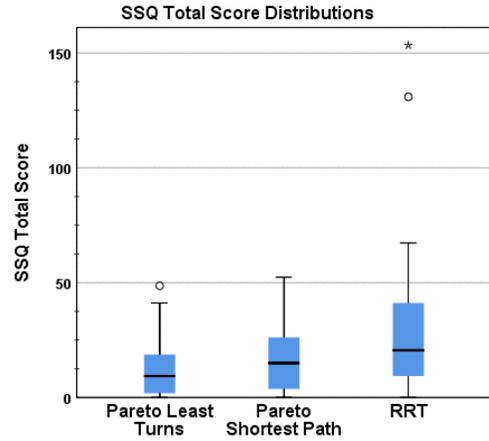

Fig. 7. Comparison of total weighted SSQ scores after watching each path video. Higher scores indicate more sickness symptoms. The black lines through the center of the boxes delineate the median total scores. The circles are outliers and the star is an extreme outlier.

preference for certain visual features or depending upon the task or environment.

V. CONCLUSION

In the present work, the formal definition of the framework of the human perception-optimized planning was provided, based on a multi objective optimization formulation. That framework is general in the sense that it allows modelling of motion planning problems where the human user is a key element within a robotic system, guaranteeing important aspects such as the users comfort. The framework was further illustrated by making use of the task of VR telepresence. That task is modelled to guarantee users comfort and performance of key components of the system, for example, the 360 camera performance. Solution trajectories were computed with a Pareto variant of the A* algorithm. Those Pareto solutions were compared against a trajectory obtained with a standard motion planning technique, which does not focus on guaranteeing any user oriented aspect.

Through experimentation on human subjects, it was validated that a solution designed to address the human perception-optimized planning problem can result in trajectories that are more comfortable for the users, or trajectories that they might prefer for other reasons. Even though the human subjects experiments were carried out in a simulated VR telepresence system, we believe that our findings justify human perception-optimized planning and apply to VR telepresence. Experimentation in VR-based telepresence systems with physical robots and 360 cameras is left for future work.

Besides user studies with a real robot, there are many other research directions to explore. First, the user's method of choosing the destination (such as a minimap, or choosing a point in the field of view), must be researched. Then, there are multiple potential criteria that should be studied. Consider user's wayfinding capabilities: for example, should other rotation angles besides multiples of 45 degrees be allowed through another planner, or can we show through

results, we can see that viewing the Pareto shortest path did not result in a statistically significant increase in SSQ total weighted scores from the baseline to the post-test, though the increase was nearly significant. The RRT, however, did result in a statistically highly significant increase in SSQ scores from the baseline to the post-test. As larger scores indicate that more sickness symptoms were experienced, these tests confirm that the Pareto shortest path was the second most comfortable and the RRT was the least comfortable. The comfort comparison questionnaire (Fig. 5) also supports that trend. The Pareto least turns path was selected by the most number of subjects as the most comfortable, followed by the Pareto shortest path, and finally the RRT path.

Surprisingly, the preference answers (Fig. 6) provide a different ranking. The Pareto shortest path was the most preferable, followed by the Pareto least turns, and the RRT path last. This is an interesting finding worthy of further investigation: that users may not always prefer the most comfortable trajectories. This might be related to the same issue pointed out above; the final part of the Pareto least turns path passes through a hallway, which could be perceived by the users as less interesting compared to the other paths that traversed through a gallery with sculptures. Because this occurred despite the fact that subjects were explicitly told not to pay attention to the pieces of art, additional criteria may need to be added in the cost functional to evaluate the user's

experiments the intuition that the current planner helps users retain their sense of direction? Additionally, users often look for certain objects in an environment, for example pieces of art in the museum, faces in a cocktail party, or exit signs at an airport. If there is a criterion that would allow users to see more of this sort of important objects, it would be a valuable finding. Additionally, can we predict the weighting and prioritizing of criteria from prior information of users, such as age, event type or gaming experience? This could be done using machine learning techniques on data obtained through human subjects experimentation; the users could then be allowed the final adjustment of weights themselves, or they could be learned even further through the users' gaze direction. Finally, we are interested in finding additional means to reduce the experienced VR sickness, such as compensating for the 360 camera motion by adding degrees of freedom to it and thus "unwinding" the rotations (the user would not experience rotation even when the robot's base is rotating), or even use time scaling on the trajectory to transform the visual stimuli to a comfortable one.

REFERENCES

[1] P. Boissy, H. Corriveau, F. Michaud, D. Labonté, and M.-P. Royer. A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects. *Journal of telemedicine and telecare*, 13(2):79–84, 2007.

[2] B. Bowins. Motion sickness: A negative reinforcement model. *Brain research bulletin*, 81(1):7–11, 2010.

[3] A. Chinchuluun, P. M. Pardalos, A. Migdalas, and L. Pitsoulis. *Pareto optimality, game theory and equilibria*. Springer, 2008.

[4] C. A. Claremont. The psychology of seasickness. *Psyche*, 11:86–90, 1931.

[5] S. V. G. Cobb, S. Nichols, A. Ramsey, and J. R. Wilson. Virtual reality-induced symptoms and effects (vrise). *Presence: Teleoperators & Virtual Environments*, 8(2):169–186, 1999.

[6] N. I. Durlach and A. S. Mavor. *Virtual reality: scientific and technological challenges*. National Academies Press, 1995.

[7] S. Hashizume, I. Suzuki, K. Takazawa, R. Sasaki, and Y. Ochiai. Telewheelchair: the remote controllable electric wheelchair system combined human and machine intelligence. In *Proceedings of the 9th Augmented Human International Conference*, page 7. ACM, 2018.

[8] Y. Heshmat, B. Jones, X. Xiong, C. Neustaedter, A. Tang, B. E. Riecke, and L. Yang. Geocaching with a beam: Shared outdoor activities through a telepresence robot with 360 degree viewing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018.

[9] S. Hughes, J. Manojlovich, M. Lewis, and J. Gennari. Camera control and decoupled motion for teleoperation. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)*, volume 2, pages 1339–1344. IEEE, 2003.

[10] A. Kemeny, P. George, F. Mérienne, and F. Colombet. New VR navigation techniques to reduce cybersickness. *Electronic Imaging*, 2017(3):48–53, 2017.

[11] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[12] S. Kratz and F. R. Ferriera. Immersed remotely: Evaluating the use of head mounted devices for remote collaboration in robotic telepresence. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 638–645. IEEE, 2016.

[13] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch. Human-aware robot navigation: A survey. *Robotics and Autonomous Systems*, 61(12):1726–1743, 2013.

[14] S. M. LaValle. *Virtual reality*. Cambridge University Press, 2020.

[15] S. M. LaValle and S. A. Hutchinson. Optimal motion planning for multiple robots having independent goals. *IEEE Trans. on Robotics and Automation*, 14(6):912–925, December 1998.

[16] S. M. LaValle and J. J. Kuffner Jr. Randomized kinodynamic planning. *The international journal of robotics research*, 20(5):378–400, 2001.

[17] D. A. Lazewatsky and W. D. Smart. An inexpensive robot platform for teleoperation and experimentation. In *2011 IEEE International Conference on Robotics and Automation*, pages 1211–1216. IEEE, 2011.

[18] M. K. Lee and L. Takayama. Now, i have a body: Uses and social norms for mobile remote presence in the workplace. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 33–42. ACM, 2011.

[19] L. Mandow and J. L. Pérez De la Cruz. A new approach to multiobjective A* search. In *IJCAI*, volume 8. Citeseer, 2005.

[20] H. Martins and R. Ventura. Immersive 3-D teleoperation of a search and rescue robot using a head-mounted display. In *2009 IEEE Conference on Emerging Technologies & Factory Automation*, pages 1–8. IEEE, 2009.

[21] M. Minsky. Telepresence. *Omni Magazine*, 38(4):217230Murray, 1980.

[22] S. Nichols and H. Patel. Health and safety implications of virtual reality: a review of empirical evidence. *Applied ergonomics*, 33(3):251–271, 2002.

[23] Y. Oh, R. Parasuraman, T. McGraw, and B.-C. Min. 360 VR based robot teleoperation interface for virtual tour. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*, 2018.

[24] E. Paulos and J. Canny. Social tele-embodiment: Understanding presence. *Autonomous Robots*, 11(1):87–95, 2001.

[25] I. Rae, B. Mutlu, and L. Takayama. Bodies in motion: mobility, presence, and task awareness in telepresence. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2153–2162. ACM, 2014.

[26] J. T. Reason and J. J. Brand. *Motion sickness*. Academic press, 1975.

[27] L. Rebenitsch and C. Owen. Individual variation in susceptibility to cybersickness. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 309–317, 2014.

[28] W. Robinett. Synthetic experience: A proposed taxonomy. *Presence: Teleoperators & Virtual Environments*, 1(2):229–247, 1992.

[29] A. H. Rupert. Motion sickness etiology: An alternative to treismans evolutionary theory. In *Spatial Orientation Symposium in Honor of Fred Guedry, Institute of Human and Machine Cognition, Pensacola, FL*, 2010.

[30] M. V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339, 2005.

[31] B. S. Stewart and C. C. White III. Multiobjective A*. *Journal of the ACM (JACM)*, 38(4):775–814, 1991.

[32] B. Stoll, S. Reig, L. He, I. Kaplan, M. F. Jung, and S. R. Fussell. Wait, can you move the robot? examining telepresence robot use in collaborative teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 14–22, 2018.

[33] P. Stotko, S. Krumpen, M. Schwarz, C. Lenz, S. Behnke, R. Klein, and M. Weinmann. A VR system for immersive teleoperation and live exploration with a mobile robot. *arXiv preprint arXiv:1908.02949*, 2019.

[34] M. Treisman. Motion sickness: an evolutionary hypothesis. *Science*, 197(4302):493–495, 1977.

[35] S. Weech, S. Kenny, and M. Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology*, 10:158, 2019.

[36] C. Widdowson, I. Becerra, C. Merrill, R. F. Wang, and S. LaValle. Assessing postural instability and cybersickness through linear and angular displacement. *Human factors*, 2019.

[37] D. Yi, M. A. Goodrich, and K. D. Seppi. MORRF*: Sampling-based multi-objective motion planning. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[38] J. Yu and S. M. LaValle. Structure and intractability of optimal multi-robot path planning on graphs. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[39] J. Zhang, E. Langbehn, D. Krupke, N. Katzakis, and F. Steinicke. Detection thresholds for rotation and translation gains in 360 video-based telepresence systems. *IEEE transactions on visualization and computer graphics*, 24(4):1671–1680, 2018.