Natural Motion Planning for Visually-guided Tasks in Representation Space

Zhenzhen Xiang and Jianbo Su Department of Automation Shanghai Jiao Tong University Shanghai, China zzxiang.sjtu@gmail.com

Abstract—This paper proposes a Hierarchical Task-guided Motion Planning (HTMP) scheme integrating Representation Space (R-space) model to address the motion planning problem in complex visually-guided tasks. The main characteristic of HTMP scheme is the interaction between task level and motion level. Such an interaction is based on the modification of R-space. When current R-space model and planner is unable to generate a feasible path, the dimension of R-space can be expanded to increase the flexibility of planning, which meets the extra requirements of natural motion. Simulation results verified the performance of the proposed method in various tasks fulfilled by a humanoid robot.

Index Terms—natural motion planning, visually-guided task, representation space, humanoid robot

I. INTRODUCTION

In the past few decades, mobile robot, especially the humanoid robot, has been devoted great efforts due to its potential of accomplishing complex tasks in daily life. In order to endow the robot with the ability of sensing and perceiving the environment, many sensors are mounted on the robot. Among these various types of sensors, vision sensor has attracted a lot attention because its input contains rich information in a relatively fast speed. A lot of research has been focused on how to plan the robot's motion and behavior with the guide of its visual information, which leads to a booming field of visually-guided motion planning [1] and navigation [2].

A household application is shown in Fig. 1 as an example. In such a scenario, a humanoid robot NAO [3] with a camera on its head is required to deliver an object, e.g., a can of coke, to a certain person. To accomplish this task, the robot needs to grasp the coke with the help of its vision sensor, then searches and recognizes its target person, and finally plans its path and approaches the target to finish the task. During the process above, on the one hand, the robot needs to plan a path to the target while avoiding collision with obstacles, which is a traditional motion planning problem; however, on the other hand, when doing the planning, the target should be maintained in the robot's field of view all the time to ensure a continuous image input for target monitoring, tracking and identifying. Moreover, the robot may be requested to finish the task more naturally, which means, for example, it should orient its body to the path when turning its head to search for a target, and



Fig. 1. An example of a visually-guided task for a humanoid robot NAO to accomplish in household applications visualized in Webots

align its body with the field of view when passing the coke in order to make its intention easily understood by the person. Obviously the latter is no longer a traditional planning problem but a new challenging one that needs to find a more natural path under various types of constraints caused by the task, environment, as well as the robot itself, which still remains an open problem.

To address the problem above, different attributes of states (e.g., the position and pose of the robot, the orientation of the camera) should be considered at the same time in the planning process in order to achieve the global optimal path. However, from the view of traditional motion planning, the planning is generally completed either in the task space (T-space) [4] or in the configuration space (C-space) [5]. The T-space planning considers the robot's state in its workspace, e.g., the position and pose of the robot in previous case, while



Fig. 2. The hierarchical task-guided motion planning scheme

the C-space planning takes place in the joint level, e.g., the rotation angle of the robot's head in previous case (which also determines the orientation of the camera). For visually-guided tasks which adopt the traditional motion planning algorithms, one important part of related work is maintaining the visibility of a moving target [6]–[8]. In these researches, a mobile robot must maintain a moving target in its field of view and avoid collision and/or occlusion caused by obstacles in the meanwhile. However, since the camera is usually fixed to the mobile robot in such cases, the motion planning strategies are only considered in T-space. Therefore, if traditional motion planning algorithms are applied to the aforementioned problem in Fig. 1, only the path trajectory in T-space can be derived, which apparently cannot fulfill the natural motion requirements as we described before.

Comparing with the previous work based on traditional motion planning methods, another part of researches about visually-guided motion planning focus on incorporating path planning in visual servoing. The main idea for path planning in visual servoing is to generate a feasible image feature trajectory with respect to image and robot constraints and then to servo the robot to follow the planned path [9]. A large number of techniques have been developed to deal with the relationship between camera space (T-space) and robot's joint space (C-space), which goes beyond the ability of traditional motion planning methods. In [10], an Alternate Task space And Configuration space Exploration (ATACE) framework was first proposed, and then was applied to the path planning for imagebased control of wheel mobile manipulators [11]. Although various attributes of states and constraints can be taken into account under the ATACE framework, alternatively exploring the T-space and C-space is just capable of achieving a feasible path instead of a natural one.

In our previous work [12], a general motion planning strategy, i.e., Representation Space (R-space) based framework, was proposed and applied to the analysis of whether the task can be realized by the robot. The basic idea of Rspace is to construct a universal space containing all different attributes of states which are used to represent a task, such as the states in T-space, C-space, image space (I-space) or even the constraints of the environment and task, etc. Then, the realizability of a task is equivalent to whether a collision-free path from initial state to final state can be found. If the task is realizable, then a global optimal path can be obtained by using traditional optimal motion planners; or, the unrealizable tasks can turn into realizable ones by reconfiguring the robot and/or modifying system/task constraints.

In this paper, we extend our efforts on planning an optimal path for a single goal in R-space to further find more natural paths in a complex task. We devised a Hierarchical Taskguided Motion Planning (HTMP) scheme (see Fig. 2) based on R-space to fulfill the requirements of natural motion. The main advantage of HTMP scheme is the interaction between task level and motion level. Such an interaction is based on the dimension expansion of R-space when facing a unfeasible task, which will iteratively meet the extra requirements of natural motion. Thus optimal and natural paths for the task can be achieved by the optimization among various attributes of states.

The proposed HTMP scheme was applied to generate more natural paths in visually-guided motion planning tasks. Section II presents the procedure of constructing the visually-guided motion model in R-space, which is the basis of the HTMP scheme. In Section III, a RRT*-based motion planner is given, and then we associate natural motion requirement with a set of constraints including robot constraints, environment constraints and visibility constraints, which leads to different natural motion modes. In Section IV, comprehensive simulation results in both simple and complex scenarios with various task goals are given to verify the effectiveness of the proposed method, followed by the conclusions in Section V.

II. VISUALLY-GUIDED TASK MODEL IN REPRESENTATION SPACE

A. Preliminaries for the Visually-guided Task

We assume that a humanoid robot (with a camera on its head) needs to perform a visually-guided task (e.g., deliver a can of coke to a specific person as we mentioned in Section I) in a living room scenario. Fig. 3 gives a more accurate and detailed description to the task. To fulfill the task requirements, the robot should plan its path under different constraints, including robot/physical constraints, obstacle/environment constraints, and visibility/task constraints.



Fig. 3. A humanoid robot performing the visually-guided task

1) Robot constraints: As Fig. 3 shows, we assume that the humanoid robot moves on a planar surface. A built-in camera is on its head with two rotational degree of freedom, yaw and pitch. Point O is the origin of the fixed frame F_O . The position of the robot in F_O is $N(x_N, y_N)$, and the pose of the robot, i.e., the angle rotated from x-axis to the orientation of the robot, is θ , where $\theta \in [0, 2\pi)$. The state of the robot in the plane (or T-space) is defined as the combination of its position and pose in F_O :

$$\boldsymbol{Z}_T = (x_N, y_N, \theta) \subseteq \mathbb{R}^3. \tag{1}$$

As the robot's head can rotate independently to its body in a certain range, we also define the yaw angle rotated from the orientation of its body as α , where $\alpha \in [-\alpha_m, \alpha_m]$, and the pitch angle as γ , where $\gamma \in [-\gamma_m, \gamma_m]$. Thus the C-space of the robot can be constructed as:

$$\mathbf{Z}_C = (\alpha, \gamma) \subseteq \mathbb{R}^2.$$
⁽²⁾

The horizontal and vertical field of view of the robot are denoted as φ and ψ , respectively.

2) Obstacle constraints: In the task scenario, various kinds of objects such as furniture and household appliances can be considered as obstacles, which should be taken into account when the robot plans its path. Every obstacle (e.g., the orange cuboid in Fig. 3) is defined with its position and contour in F_O for collision detection. For simplicity, we assume the obstacles can just cause physical collision, and no occlusion would take place, since a human is usually much easier to be observed in a room scenario.

3) Visibility constraints: The primary requirement of the aforementioned visually-guided task is to maintain the target in the field of view when the robot is approaching to it. Given the state of the robot $z_N(x_N, y_N, \theta, h)$ and the state of the target person $z_T(x_T, y_T, H)$, we use β and ζ to denote the angle of the head of the target person in the robot's field of view, then if β and ζ satisfy:

$$|\beta| \le \frac{\varphi}{2}, \quad |\zeta| \le \frac{\psi}{2},\tag{3}$$

then the target person is visible for the robot; or, the target person is invisible.

B. Modeling in Representation Space

It is obvious that both T-space states (i.e., the position and pose of the robot) and C-space state (i.e., the yaw and pitch angle of the robot's head) are included in the visually-guided task, which exceeds the capability of traditional motion planning frameworks. Therefore, we can construct the following model in R-space which combines the T-space and C-space, so as to plan and optimize the robot's path under all of the constraints:

$$\boldsymbol{R}_N = (x_N, y_N, \theta, \alpha, \gamma) \subseteq \mathbb{R}^5.$$
(4)

With the model in R-space, we can still utilize traditional optimal motion planning algorithms, yet achieve various optimal paths for multiple task goals, which will be described in detail in Section III.

III. MOTION PLANNING AND OPTIMIZATION FOR NATURAL PATHS

A. Motion Planning with RRT*

Suppose the R-space \mathbf{R}_N has been constructed. Let $\mathbf{R}_{N_{obs}}$ denote the collision subspace in \mathbf{R}_N , and $\mathbf{R}_{N_{free}} = \mathbf{R}_N \setminus \mathbf{R}_{N_{obs}}$ denotes the collision-free subspace. Given a cost function c(r), the optimal motion planning problem is to find a feasible path $r(t) \in \mathbf{R}_{N_{free}}$ for $t \in [0, t_f]$ from an initial state $r(0) = r_{\text{int}}$ to the goal region $r(t_f) \in \mathbf{R}_{goal}$ and minimize the cost function.

Algorithm 1: $\tau = (V, E) \leftarrow RRT^*(r_{int})$ $\tau \leftarrow InitializeTree();$ 2 $\tau \leftarrow InsertNode(\emptyset, r_{int}, \tau);$ **3** for i = 1 : N do $r_{rand} \leftarrow Sample(i);$ 4 $r_{\text{nearest}} \leftarrow Nearest(\tau, r_{\text{rand}});$ 5 $(r_{\text{new}}, u_{\text{new}}) \leftarrow Steer(r_{\text{nearest}}, r_{\text{rand}});$ 6 7 if $ObstacleFree(r_{new})$ then 8 $R_{\text{near}} \leftarrow Near(\tau, r_{\text{new}}, |V|);$ $r_{\min} \leftarrow ChooseParent(R_{near}, r_{nearest}, r_{new});$ 9 $\tau \leftarrow InsertNode(r_{\min}, r_{\text{new}}, \tau);$ 10 $\tau \leftarrow ReWire(\tau, R_{near}, r_{min}, r_{new});$ 11 end 12 13 end 14 return τ ;

Many traditional motion planning algorithms can be adopted to solve the problem above. Since the incremental samplingbased algorithms such as the Probabilistic RoadMap (PRM*) and the Rapidly-exploring Random Tree (RRT*) [13] are proven to be probabilistically complete, asymptotically optimal and computationally efficient, we choose the RRT* algorithm as the motion planner to generate the optimal path in $R_{N_{free}}$.

The pseudocode of the RRT* algorithm is given in Algorithm 1. To solve the optimal motion planning problem, the RRT* algorithm builds and maintains a tree $\tau = (V, E)$ in $\mathbf{R}_{N_{free}}$, which consists of a set of vertices $V \in \mathbf{R}_{N_{free}}$ and edges $E \subseteq V \times V$. The RRT* algorithm iteratively uses a group of basic procedures as shown in the Algorithm 1 from line 4 to line11. The detailed meaning and computational process of each function can be found in [13]. Finally the RRT* algorithm returns the optimal path that minimizes the cost function.

B. Optimization for Natural Paths

To obtain a desired natural path using the RRT* algorithm, the cost function as well as the constraints should be given properly, especially for motion planning problems defined in R-space. As we mentioned in Section I, one of the characteristics of R-space, i.e., it can involve various attributes of states and even constraints, make it possible to generate different natural paths in a complex task. Therefore, in the context of accomplishing the visually-guided task in household applications, two different natural motion modes can be derived by the humanoid robot according to recent vision research [14].

1) Body orientation sticking to the path: This mode reflects one type of natural motion requirement for the robot in the visually-guided task, especially under the circumstances of searching and recognizing the target. When the robot is searching and recognizing the specific person which is required to approach, to rotate its head rather than to adjust its body's orientation is much more energy efficient and less time consuming. Thus the orientation of the robot's body should stick to the path within an acceptable deviation. In order to control the deviation between the body orientation and the path, the basic idea is to consider the deviation of every new node that is about to be inserted to the tree or rewired to another node, as Fig. 4 shows. The arrows represent different body orientations of the robot, and the nodes denote its positions. Then we can define a maximum path-deviation angle δ_m and detect whether the path-deviation angle δ of the new node satisfies:

$$|\delta| \le \delta_m. \tag{5}$$

If the condition is not satisfied, we consider it as a path collision, then the node will be discarded and the RRT* algorithm continues the next iteration.



Fig. 4. The path-deviation angle of a new node

2) Body orientation sticking to the field of view: This mode is another type of natural motion requirement for the robot especially in the situation of passing the object to the specific person. Under such a circumstance, the robot needs to adjust its body to the same orientation as its field of view, i.e., the orientation of its head, so that the person can easily notice the intention of the robot and prepare for catching the object.

To achieve the goal above, we can select a proper value for α_m to control the deviation between the orientation of the robot's body and the center of its horizontal field of view. For example, if we want its body exactly align with its field of view, α_m can even be set to zero.

IV. SIMULATION RESULTS

Two groups of simulation are designed to evaluate the performance of the proposed R-space based motion planning algorithms in solving the visually-guided task: one is the 3-D case, and the other is the 5-D case. In 3-D case, the motion planning is accomplished in the space constructed by humanoid's position and posture. This space can be regarded as a special case of R-space. In 5-D case, we consider both the humanoid's posture and its camera orientation in the visually-guided task. In both cases, the visibility constraints are considered as task constraints to ensure that the target was always in humanoid's vision.

TABLE ICONFIGURATIONS OF THE ROBOT

Robot size			Robot vision			
<i>L</i> (m)	<i>W</i> (m)	$H(\mathbf{m})$	α_m (°)	$\varphi\left(^{\circ} ight)$	$\gamma_{m}\left(^{\circ} ight)$	ψ (°)
0.3	0.2	0.5	90	60	30	40



Fig. 5. Simulation scenarios for visually-guided tasks

TABLE II SAMPLING INTERVALS IN R-SPACE FOR RRT*

$\Delta x(m)$	$\Delta y(m)$	$\Delta \theta (\circ)$	$\Delta \alpha (\circ)$	$\Delta\gamma\left(^{\circ} ight)$
0.04	0.04	10	5	15

The configurations of the humanoid robot in the following simulation are shown in Table I, which is in accordance with a NAO robot. Two simulation scenarios are shown in Fig. 5 as the approximation of the NAO robot in indoor environments with different complexity. In the simple scenario, the performance of R-space based planning algorithms are respectively tested under different goals under the same condition; in the complex scenario, a complete visually-guided task is given to the robot in order to evaluate the overall performance of the algorithms. In Fig. 5, the orange rectangle represents the robot; the blue arrow represents its body orientation; the blue triangle denotes its view; the circle is the target region. The sample intervals of different variables in R-space for RRT* algorithm is given in Table II, which can be adjusted under the consideration of both accuracy and computational efficiency. In addition, the number of iterations for RRT* algorithm is set to $N_{5D} = 200,000$ in both 3-D and 5-D case. Finally, a same random sampling sequence is adopted by the Sample() function in Algorithm 1 so as to compare the performance of different cases.



Fig. 6. Optimal path generated in 3-D case (path length: 2.22 m)

A. Simple Scenario Case with 3-D R-space

In this case, following the procedures given in Section II and Section III, first the task is modeled in R-space, and then two optimal paths for different goals are generated by exploring the collision-free subspace $R_{N_{free}}$ with RRT* algorithm.

1) R-space model: In 3-D case, the R-space is defined as

$$\boldsymbol{R}_{N_{3D}} = (x_N, y_N, \theta) \subseteq \mathbb{R}^3.$$
(6)

According to the scenario defined at the beginning of this section, the ranges of different states in $\mathbf{R}_{N_{3D}}$ are: $x_N \in [0, 3]$, $y_N \in [0, 2]$ and $\theta \in [0, 2\pi)$, respectively, and then the R-space can be gridded with sampling intervals in Table II. For each sampling point in R-space, we can determine whether it collides with any constraints so as to divide the R-space into two subspaces, i.e., collision subspace and collision-free subspace.

2) Planning with RRT*: After the construction of R-space model, we can apply RRT* algorithm to explore in the collision-free subspace. The simulation results are shown in Fig. 6. Here the trees generated by RRT* algorithm is plotted in green lines. We can find that although the humanoid robot achieves the shortest path, the poses of its body and head are only decided by the obstacles and the target. Hence the humanoid robot will not be considered as a agent with natural motion in a social perspective.

B. Simple Scenario Case with 5-D R-space

In this case, since the construction process of the model in R-space is similar as that in 3-D case, which is shown as:

$$\boldsymbol{R}_{N_{5D}} = (x_N, y_N, \theta, \alpha, \gamma) \subseteq \mathbb{R}^5, \tag{7}$$

we focus our attention on the optimization of different natural paths in the visually-guided task. As described in Section III-B, the robot can obtain two types of natural paths by applying RRT* algorithm.

- Body orientation sticking to the path: To constrain the maximum deviation of the robot's body orientation from the path, we set $\delta_m = \pi/3$ under the constraint of visibility.
- *Body orientation sticking to the field of view:* In order to make the robot face to its vision, we additionally restrict



(a) Body orientation sticking to the path (path length: 2.38 m)



(b) Body orientation sticking to the field of view (path length: 2.36 m)

Fig. 7. Optimal paths for different goals in 5-D case

the range of its head rotation angle α to $[-\pi/6, \pi/6]$ as a task constraint, although the robot is able to exceed this range in practical.

Fig. 7(a) and 7(b) show the optimal paths for the abovementioned two kinds of natural paths, respectively. We can find that the results are approximate to those in the 3-D case. However, such results are more satisfactory than the one in Fig. 6 due to the dimensional expansion of R-space.

In Fig. 7(a), the humanoid robot keeps the orientation of its body to the path while allowing its head to turn a large angle to maintain the target in the field of view, especially in the beginning of its path. But with such a requirement it has to choose a further path to avoid collision with obstacles, comparing with the path in Fig. 6. It should also be noted that the number of nodes in the tree decreases dramatically due to the strict deviation constraints. This suggests that a much larger number of iterations is required in RRT* algorithm to further optimize the path.

In Fig. 7(b), the robot manages to orient its body to the direction of vision when approaching the target region, where the body of the robot always faces to the right guided by the target. Though the energy cost increases comparing to the optimal path in Fig. 6, it actually behaves in a more natural way in most of time, thus it is easy for the human to see its face and understand its intention, and then prepare for catching



Fig. 8. The optimal path in complex scenario: (1) $0 \rightarrow 1$ is optimized as the same goal in 3-D case; (2) $1 \rightarrow 2 \& 2 \rightarrow 3$ are both optimized for the natural mode of "body orientation sticking to the path"; (3) $3 \rightarrow 4$ is optimized for the natural mode of "body orientation sticking to the field of view"

the object in its hand.

C. Complex Scenario Case with 5-D R-space

In this scenario, we follow the discussion of the visuallyguided task described in Section I, and try to solve such a problem with the proposed R-space based methods for natural motion. Suppose the places for the humanoid robot to explore have been decided, and task flow and requirements are given as follows:

- 0 → 1: The humanoid robot needs to find the shortest path from its initial position to region 1 and get a can of coke from the fridge. When approaching the target region, the robot should maintain it in its field of view and keep monitoring, so that if other people are using the fridge it can stop before the target region and wait for a moment.
- 1 → 2: After getting the coke in hand, the humanoid robot should explore the room and search for the target person (which is denoted as a yellow rectangle in Fig. 9). When walking along the path, the humanoid robot is expected to keep its body orientation sticking to the path and search for its target person mainly by rotating its head. In region 2, two people (green and pink rectangles) can be observed and identified by the identification technologies, e.g., face recognition, within a certain distance (inside both the green and pink circle), then the humanoid robot will find that neither of them is its target person.
- 2 → 3: The humanoid robot needs to continue its searching in the room, and then it will successfully find its target person after arriving at the region 3 inside the yellow circle.
- 3 → 4: In order to pass the coke to its target person, the humanoid robot should approach to the region 4 slowly, and align its body and head to the target person in order to make the person notice its intention and get ready for catching the coke.

According to the task requirements, we use the proposed Rspace based methods to plan and optimize the path for natural motion requirements. The optimal path for the task above is shown in Fig. 8. We can find that in each stage of the task, the path is optimized for a certain goal and the visibility of the target is always maintained in the robot's field of view, so that the basis for monitoring, searching and identifying operations can be ensured and performed with a more efficient and natural way.

V. CONCLUSIONS AND FUTURE WORK

In this paper, the motion planning problem in visuallyguided tasks performed by a humanoid robot is considered. Due to the complexity of the task, we proposed a Hierarchical Task-guided Motion Planning (HTMP) scheme based on Representation Space framework, in order to plan and optimize the motion of the humanoid robot among various attitudes of states for natural paths under different constraints. Simulation results reflect the effectiveness of the proposed method in accomplishing a visually-guided task in both simple and complex household scenarios. Future work will be focused on the efficiency of the optimal motion planner in representation space for complex visually-guided tasks and preparing for realtime applications.

ACKNOWLEDGMENT

This work was supported by National Nature Science Foundation of China Project (Grant No. 61221003).

REFERENCES

- D. Kragić, L. Petersson, H. I. Christensen. Visually guided manipulation tasks. *Robotics and Autonomous Systems*, 2002, 40(2): 193-203.
- [2] F. Bonin-Font, A. Ortiz, G. Oliver. Visual navigation for mobile robots: A survey. *Journal of Intelligent and Robotic Systems*, 2008, 53(3): 263-296.
- [3] D. Gouaillier, V. Hugel, P. Blazevic, et al. Mechatronic design of NAO humanoid. *IEEE International Conference on Robotics and Automation*, 2009: 769-774.
- [4] M. Behnisch, R. Haschke, M. Gienger. Task space motion planning using reactive control. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010: 5934-5940.
- [5] L. Jaillet, J. Corts, T. Simon. Sampling-based path planning on configuration-space costmaps. *IEEE Transactions on Robotics*, 2010, 26(4): 635-646.
- [6] S. M. LaValle, H. H. Gonzalez-Banos, C. Becker, et al. Motion strategies for maintaining visibility of a moving target. *IEEE International Conference on Robotics and Automation*, 1997, 1: 731-736.
- [7] T. Muppirala, S. Hutchinson, R. Murrieta-Cid. Optimal motion strategies based on critical events to maintain visibility of a moving target. *IEEE International Conference on Robotics and Automation*, 2005: 3826-3831.
- [8] N.R. Gans, G. Hu, K. Nagarajan, et al. Keeping multiple moving targets in the field of view of a mobile camera. *IEEE Transactions on Robotics*, 2011, 27(4): 822-828.
- [9] M. Kazemi, K. Gupta, M. Mehrandezh. Path-planning for visual servoing: A review and issues. *Visual Servoing via Advanced Numerical Methods*. Springer London, 2010: 189-207.
- [10] Z. Yao, K. Gupta. Path planning with general end-effector constraints. *Robotics and Autonomous Systems*, 2007, 55(4): 316-327.
- [11] M. Kazemi, K. Gupta, M. Mehrandezh. Path planning for imagebased control of wheeled mobile manipulators. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012: 5306-5312.
- [12] J. Su, W. Xie. Motion planning and coordination for robot systems based on representation space. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B: Cybernetics, 2011, 41(1): 248-259.
- [13] S. Karaman, E. Frazzoli. Sampling-based algorithms for optimal motion planning. *The International Journal of Robotics Research*, 2011, 30(7): 846-894.
- [14] J. M. Franchak, K. E. Adolph. Visually guided navigation: Headmounted eye-tracking of natural locomotion in children and adults. *Vision research*, 2010, 50(24): 2766-2774.