Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2023) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

5th International Conference on Industry 4.0 and Smart Manufacturing

Applying grid world based reinforcement learning to real world collaborative transport

Alexander Hämmerle[a,*], Christoph Heindl[a], Gernot Stübl[a], Jenish Thapa[a], Edoardo Lamon[b,c], Andreas Pichler[a]

[a] Profactor GmbH, Im Stadtgut D1, 4407 Steyr-Gleink, Austria
[b] Human-Robot Interfaces and Interaction, Istituto Italiano di Tecnologia, Via S. Quirico 19d, 16163 Genoa, Italy
[c] Dept. of Information Engineering and Computer Science, Università di Trento Via Sommarive, 9, Povo, Italy

Abstract

Motivated by transportation tasks on construction sites, this contribution deals with an AI-driven approach to human-robot collaborative transportation. An essential part of the considered problem is navigating the robot to the object to be transported, in the presence of other moving items like the human moving to the object. Robot navigation is tackled with reinforcement learning, and the impact of randomness in the other moving items' behaviour on the robot's training performance is investigated. Results show that the move failure rate of the trained robot policy increases, when the behavioural patterns in the human's movements are disturbed by randomness. On the other hand, when both human and robot are connected to the object, the navigation problem is delegated to the human, which guides the compound human-object-robot to the goal location.

Keywords: collaborative transportation; reinforcement learning; robot navigation;

1. Introduction

In this paper, human-robot collaboration for the transportation of heavy and bulky objects at construction sites is investigated. For a human-robot collaboration, construction sites pose a challenge due to them being unstructured and dynamic environments. For the movement of heavy objects, a mixed human-robot team has to navigate uncertain and constantly changing terrain. Also, in human-robot collaborative transportation, the robot is confronted with a situation that there is no guarantee that the human involved in such a collaboration may always act strictly rational and, for example, follow a shortest path to a destination where these heavy objects are transported to. In this paper, training robot navigation policies in the presence of a human as the collaboration partner, as well as random events occurring

[*] Corresponding author. Tel.: +43(0)7252-885-160 ; fax: +43(0)7252-885-101.
E-mail address: alexander.haemmerle@profactor.at

in the environment, such as moving obstacles (workers, construction machines) are considered. These moving items may behave randomly to some extent, and the impact of that randomness on the robot's training performance is investigated. The investigations were conducted in a virtual environment. The concepts presented in this paper were developed in the context of the EU-funded CONCERT project.

The remainder of the paper is structured as follows. Related work is discussed in section 2, followed by the problem description in section 3. The approach to solve the collaborative transportation problem is described in section 4. Results are presented in section 5, followed by concluding remarks and an outlook on further research in section 6.

2. Related work

Transportation tasks often require the collaboration of multiple agents for transporting heavy load. Approaches to enable multi-robot collaboration (without human intervention) are shown e.g. in [5, 20, 21]. In contrast, approaches for transportation through human-robot collaboration are presented e.g. in [1, 10, 14, 22].

With respect to grid-world-based reinforcement learning applied to robot navigation, in [18], the performance of different deep reinforcement learning algorithms is compared. The comparison is conducted with a use case where a robot has to navigate from a random starting location to a random destination, avoiding up to two randomly located obstacles. The use case is implemented as grid world. In [9], a reward shaping framework for average learning in continuing tasks is presented. The framework is evaluated with several test cases; one of the cases is a continuing grid world, where the robot has to navigate to a given goal cell, avoiding a large obstacle. [3] tackles the problem of perceptual aliasing in robot navigation, i.e., due to sensor limitations the robot sometimes is not able to distinguish between differing world states. The work uses Sutton's grid world, see [19], as well as a simple 1-D example. [17] investigates into efficient initialisation approaches for Q-learning, using a maze-like grid world as test case. The training goal is to navigate to a specified goal location. A maze-like grid world environment is also used in [4] ; the authors present a modified scheme for reinforcement learning with separated mechanisms for positive and negative rewards, respectively. In [16], the authors investigate invariance principles for stronger generalisation in reinforcement learning; the test case is implemented as grid world: a robot is located in one room of the grid world, and it must learn to navigate to a goal location in a different room. To enter the room, the robot first has to acquire a 'key' object.

In the related work discussed above, the robot is not confronted with moving items that do not always act in a completely rational fashion. We address this research gap in this paper, training robot navigation policies in the presence of moving items with partially random behaviour.

3. Problem Description

At a two-dimensional construction site, an object has to be transported to a specific goal location and two agents (human and robot) are available for the transportation task. The agents start at random locations, i.e., the first part of the transportation task for an agent is to move to the object and connect to it. Moving to the object, the robot should not collide with the human (also moving to the object).

Figure 1 in the middle depicts the technical embedding of the problem using a Behaviour Tree notation. Behaviour Trees are a common tool for human robot interaction to foster elementary skills into higher-level tasks, see [8]. However, the major drawback is its fixed sequential execution of siblings (e.g. *move left*,*move right* etc.) making it inflexible in the execution. This paper investigates on a trained node, that dynamically dispatches robot actions until the robot is connected to the object. Please note, while this paper focuses on learning in simulation only, for the execution on the robot an interpretation of the real environment as grid world is possible with an available computer vision system based on [7].

As soon as both agents are connected to the object, the human guides the compound robot-object-human to the goal location (*MoveCompound)*. This research problem, which investigates the physical coupling of the agents, has been extensively studied in the literature [1, 10] as well as within the CONCERT project [13, 14], and hence will be mentioned in this paper for the sake of completeness, to show the overall functionality of the proposed framework. In addition to the basic transportation problem described above, a problem variant is considered. In this variant, a dynamic obstacle (e.g., another human) moves on the construction site, impeding the agents' movements.

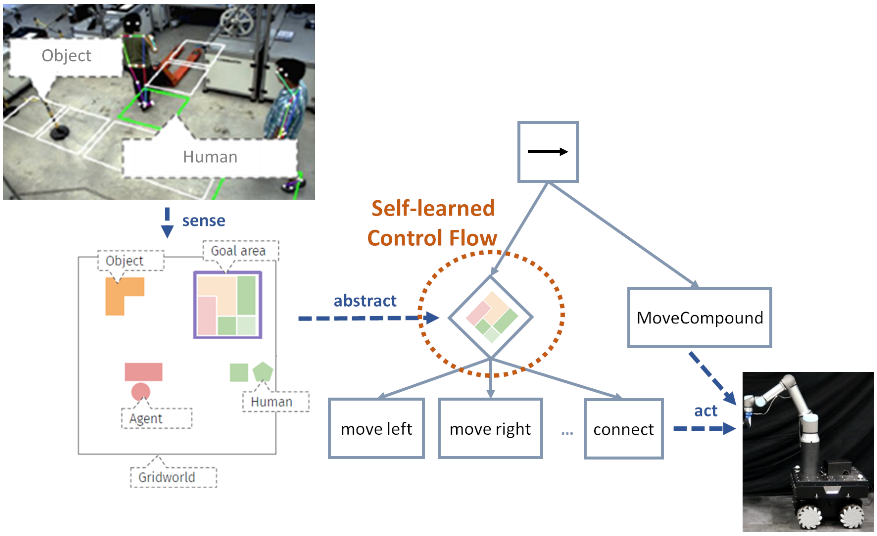

Fig. 1: Behaviour Tree notation of the transport task. The self-learned Control Flow Node dispatches the right actions sequence for connecting to the object. After connection, the *MoveCompound* node is executed. While policies are learned in simulation, a perception of the real environment as grid world is possible via a computer vision system.

4. Approach

For the first part of the transportation task, the main problem is the robot navigation behaviour. The robot should move efficiently from a random starting location to the object and connect to it. While moving, the robot should avoid collisions with the human and the moving obstacle. To implement such behaviour, deep reinforcement learning is applied.

Deep reinforcement learning

In reinforcement learning, an agent acquires decision-making skills through sequential interactions with an environment. At each time step t, the agent receives an observation s_t from the environment, and based upon s_t the agent decides on an action a_t, using a policy function $\pi(a_t|s_t)$. The training goal for the agent is to maximise the expected cumulative reward:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}, \quad \gamma \in (0, 1]. \tag{1}$$

In equation 1, γ is a discount factor, and r_t is the reward at time t, provided by the environment to the agent. The value of a policy π for a state s is defined as

$$v_\pi(s) = \mathbb{E}_\pi(R_t|s_t = s), \tag{2}$$

and the policy's action-value function is

$$q_\pi(s, a) = E_\pi(R_t|s_t = s, a_t = a), \tag{3}$$

with action a taken in state s. An optimal policy results from maximising the action-value function. In deep reinforcement learning, deep neural networks are used for the representation of policy functions, introducing the network parameter θ. With the introduction of θ it is possible to search for an optimal value for θ in the policy space $\{\pi_\theta(a_t|s_t), \theta\}$.

Policy gradient methods implement a gradient ascent approach for the optimisation of the neural network parameter θ, see [2]. A parameter update is proportional to an estimate of the objective function's gradient. A frequently used objective function in policy gradient methods is proposed in [12]:

$$L^{PG}(\theta) = \hat{\mathbb{E}}_t[\log \pi_\theta(a_t|s_t)\hat{A}_t] \tag{4}$$

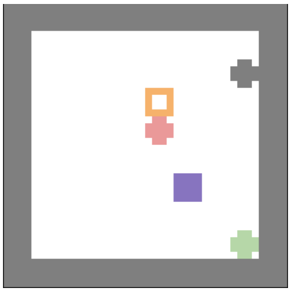

Fig. 2: Snapshot of the grid world training environment; the 10x10 grid world is surrounded by walls (grey cells); the object encoding is as follows: red cross − > robot, green cross − > human, grey cross − > moving obstacle, purple square − > goal, orange square − > object to be transported;

where $\hat{A}_t$ is an estimator of the advantage function that describes the additional benefit that could be gained by acting in the manner indicated by a_t. The combination of policy gradient approaches with action-value functions gives rise to actor-critic methods, where the critic network approximates the action-value function, and the actor network implements an approximation of the policy function. In training the networks, the critic's task is to criticise the actions taken by the actor.

In the presented work, RLlib's implementation of a Proximal Policy Optimisation algorithm (PPO) is used for reinforcement learning, see [11] and [12]. PPO algorithms belong to the family of actor-critic methods, implementing a constraint mechanism to stabilise training. Policy changes are constrained by the objective function 5, resulting in smaller increments in θ.

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)], \tag{5}$$

where $clip(r_t(\theta), 1-\epsilon, 1+\epsilon)$ clips the ratio to the interval $[1-\epsilon, 1+\epsilon]$, and $r_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t))$.

Reinforcement learning in a grid world

The training environment for the robot policy is modelled as grid world, see figure 2. Each training episode starts with random locations for robot, goal and object. The episode terminates successfully when the robot has connected to the object. In a failed episode the robot was not able to connect to the object within a given number of time steps (time out). At each time step, all dynamic items (robot, human, moving obstacle) act simultaneously. To avoid collisions, the robot has to learn to anticipate the movements of human and obstacle.

In reinforcement learning, at each time step the robot acts in the environment, and the environment's response is an observation and a reward for the action taken. Hence crucial design decisions are:

- the observation space: what is the observation that is provided by the environment to the robot?
- the action space: what are the actions that are available to the robot?
- the reward scheme: what is the reward that is provided by the environment to the robot, in response to an action?

In the presented work, the robot's observation is global, consisting of an item encoding for every cell in the grid world. The items are one-hot encoded with vector length L, hence the size of the observation space is $10x10xL$, where $10x10$ is the size of the grid world. The robot's action space is {move left, move right, move up, move down, connect to object}. To increase training efficiency, action masking is applied. In action masking, physically impossible actions are ruled out by the environment. The following action masking rules apply:

- If any item from the set {object, wall, obstacle, human} is in an adjacent cell to the robot, the action to move to that cell is ruled out.
- If the object is not in an adjacent cell to the robot, the action ”connect to object” is ruled out.

The reward is a scalar, encoding the quality of an action taken with respect to the pursued training goals. The training goals are:

1. Move to the object and connect to it;
2. Move efficiently;
3. Avoid collisions with human and obstacle;

The training goals are addressed with the following reward scheme:

1. A positive reward for successfully connecting to the object; addresses goal 1;
2. A negative reward for each action; addresses goal 2;
3. A negative reward for an illegal action, i.e., an action that resulted in a collision; addresses goal 3;

Models for human and moving obstacle

The human is modelled as an agent in the grid world with the following behaviour:

- The human moves on a shortest path to the object. Shortest paths are calculated with an A* algorithm, see [6].
- Due to the fact that the agents act simultaneously in the grid world, A* can suggest moves that would result in collisions. For example, at time T the human agent wants to move to grid cell B, but at time T the robot agent has already moved to cell B. In such a case, the move resulting from A* is ignored, and the human agent does not move.
- In addition to rationally moving along shortest paths, the human may move in a random fashion. The strength of the random component of the human agent's behaviour can be configured with parameter $p_h = [0, 1]$, where p_h is the probability that the human's move is random.

The moving obstacle's behaviour is modelled as follows:

- The moving obstacle is configured with a sequence of waypoints in the grid world. In each training episode, the obstacle starts at the first waypoint. Then, the obstacle visits the waypoints in the given sequence, moving on shortest paths. When the last waypoint has been visited, the obstacle moves on a shortest path to the first waypoint, and the movement across the sequence of waypoints is repeated. The obstacle stops when the episode terminates.
- The collision avoidance behaviour is equal to the human agent.
- Similar to the human agent, the moving obstacle can behave randomly. Parameter $p_o = [0, 1]$ configures the strength of the random component of the obstacle's behaviour, where p_o is the probability that the obstacle's move is random.

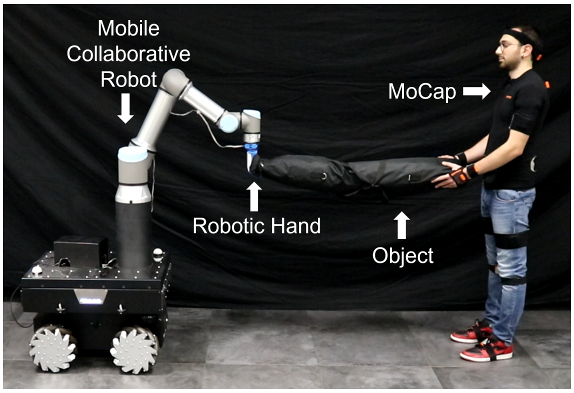

Fig. 3: Human-robot collaborative carrying of objects with unknown deformability. The framework integrates haptic information conveyed through the carried object with human kinematic information acquired from a motion capture system, generating responsive whole-body movements on a mobile collaborative robot [14].

In the remainder of the transportation task (*MoveCompound*) two problems have to be tackled: a) the robot has to assist the human in carrying the object, and b) the human has to navigate the compound human-object-robot onto the goal location.

For this manuscript, we adopt the strategy presented in [14], which consists in a novel human-robot collaborative co-carry framework characterised by 1) the possibility to transport objects with unknown deformability, by means of the combination of both force sensing and real-time human motion capture (see figure 3), and 2) obstacle-aware vibrotactile feedback during co-carry, which is motivated by the fact that occlusions of the human vision when carrying large objects may occur, compromising the environmental awareness of the human and hence demoting safety [15]. More details of the co-carry framework and results of the framework in physical settings can be found in [14, 15].

5. Results

In this section, results are presented with respect to training the robot policy in the virtual grid world. The focus is on investigating the impact of p_h and p_o on the training performance for the robot policy. The experimental setup is as follows:

- Grid world size: 10x10; the origin (0,0) is in the upper left corner;
- Episode time out: 40 time steps
- Reward for successfully connecting to the object: 1.0
- Reward for each action: -0.01
- Reward for illegal action: -0.05
- $p_h = \{0, 0.3\}$
- $p_o = \{0, 0.3\}$
- Training duration: 10 million time steps (no obstacle), 30 million time steps (moving obstacle)
- Sequence of way points for moving obstacle: (8,2), (5,2), (7,4), (2,5), (3,8)

With 50cm units in the grid world, a grid cell is 50cm x 50cm, providing enough space for a human. A grid with size 10 x 10 then covers an area of 5m x 5m, and that's what usually can be covered by a camera.

Following the specified sequence of waypoints, the moving obstacle traverses large parts of the grid world, potentially resulting in collisions with the robot. The other setup parameters were determined empirically.

Figures 4 and 5 show training results for the grid world without obstacle. The plots in figure 4 contain outliers in chart scaling; these plots are useful to get an overall impression of the training performance. Outliers are omitted in figure 5, providing more details. In each plot, the x-axis shows the number of time steps. Figure 5 clearly illustrates the impact of p_h on the training performance: with $p_h = 0.3$, the average number of illegal move actions increases, due to collisions between human and robot. Training results for the grid world with a moving obstacle are depicted in figures 6 and 7. The impact of p_h and p_o on the average number of illegal move actions is noticeable: with $p_h = 0$ and $p_o = 0$, the number of collisions is lowest. With either $p_h = 0.3$ or $p_o = 0.3$, the number of collisions increases, and with $p_h = 0.3$ and $p_o = 0.3$, the collision rate is highest.

6. Conclusion and outlook

In this contribution, an approach to human-robot collaborative transportation is presented. An essential part of the collaborative transportation problem is navigating the robot from a random initial location to the object to be transported, avoiding collisions with humans. That part is solved with an AI approach, namely reinforcement learning in a grid world. The results obtained in a virtual environment show a significant impact of the random component in the human's behaviour on the robot's training performance. If the human acts randomly to some extent, it is harder for the robot to learn to anticipate the human's movements, and thus the collision rate increases. The negative impact of randomness in the training environment on the robot's training performance is also observed in a problem variant, with a moving obstacle added to the grid world. In this case, the robot has to learn to anticipate the movements of two dynamic items (human and obstacle). Moreover, a human-robot co-carrying framework was adopted to complement

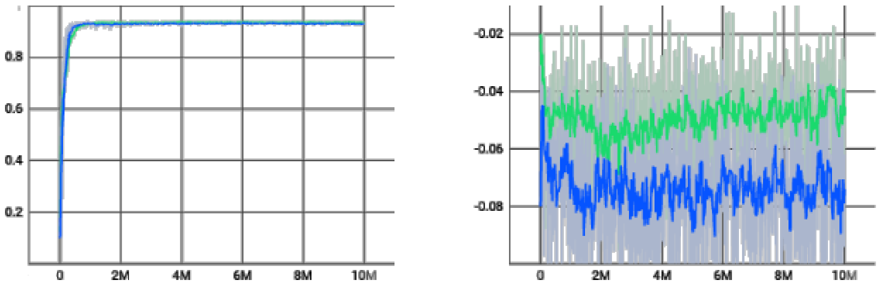

Fig. 4: Training performance for use case variant ”no obstacle”; green: $p_h = 0$, blue: $p_h = 0.3$; the left plot shows the average episode reward, the right plot shows the average number of illegal move actions per episode (an illegal move action adds -1 to the statistics).

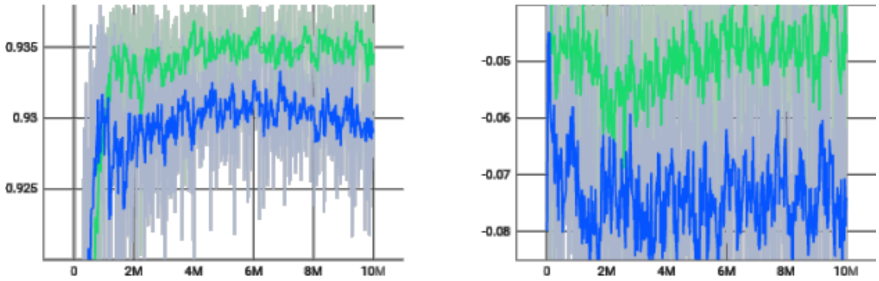

Fig. 5: Training performance for use case variant ”no obstacle”; outliers are ignored in chart scaling; green: $p_h = 0$, blue: $p_h = 0.3$; the left plot shows the average episode reward, the right plot shows the average number of illegal move actions per episode (an illegal move action adds -1 to the statistics).

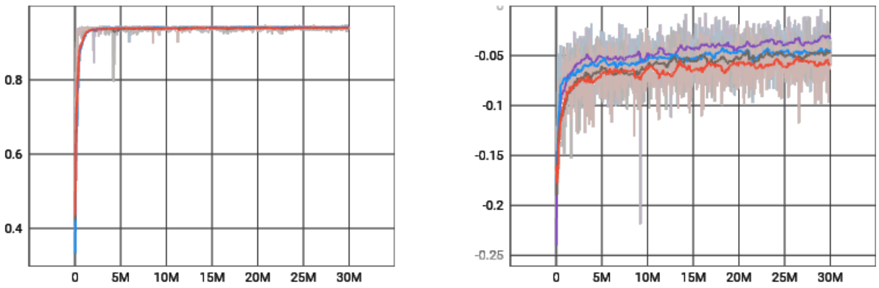

Fig. 6: Training performance for use case variant ”moving obstacle”; violet: $p_h = 0, p_o = 0$; blue: $p_h = 0, p_o = 0.3$; dark green: $p_h = 0.3, p_o = 0$; orange: $p_h = 0.3, p_o = 0.3$; the left plot shows the average episode reward, the right plot shows the average number of illegal move actions per episode (an illegal move action adds -1 to the statistics).

the navigation approach. However, further research aims at integrating the individual research results for the two parts of the collaborative transportation problem to an AI based framework for human-robot collaborative transportation.

7. Acknowledgement

This work was supported by the European Union’s Horizon 2020 research and innovation programme within the project CONCERT (grant number 101016007), and the Lighthouse project supported by the Austrian Institute of Technology (AIT). This research is also financed by research subsidies granted by the government of Upper Austria. We also acknowledge the support of the MUR PNRR project FAIR - Future AI Research (PE00000013).

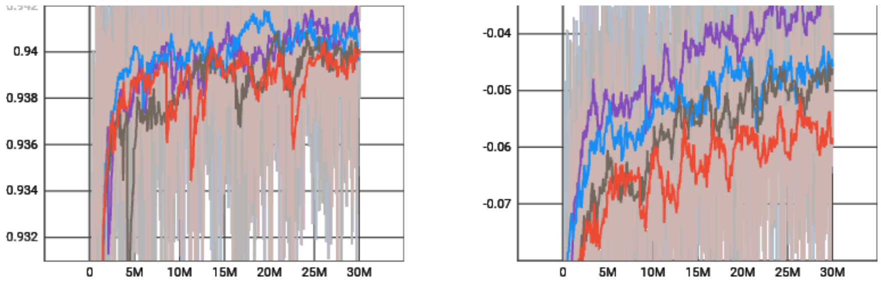

Fig. 7: Training performance for use case variant ”moving obstacle”; outliers are ignored in chart scaling; violet: $p_h = 0, p_o = 0$; blue: $p_h = 0, p_o = 0.3$; dark green: $p_h = 0.3, p_o = 0$; orange: $p_h = 0.3, p_o = 0.3$; the left plot shows the average episode reward, the right plot shows the average number of illegal move actions per episode (an illegal move action adds -1 to the statistics).

References

[1] Alevizos, K.I., Bechlioulis, C.P., Kyriakopoulos, K.J., 2020. Physical Human–Robot Cooperation Based on Robust Motion Intention Estimation. Robotica 38, 1842–1866. URL: https://www.cambridge.org/core/product/identifier/S0263574720000958/type/journal_article, doi:10.1017/S0263574720000958.

[2] Baird, L., Moore, A., 1998. Gradient Descent for General Reinforcement Learning, in: Kearns, M., Solla, S., Cohn, D. (Eds.), Advances in Neural Information Processing Systems, MIT Press. URL: https://proceedings.neurips.cc/paper/1998/file/af5afd7f7c807171981d443ad4f4f648-Paper.pdf.

[3] Crook, P., Hayes, G., 2003. Learning in a State of Confusion: Perceptual Aliasing in Grid World Navigation, in: Proceedings of Towards Intelligent Mobile Robots (TIMR 2003).

[4] Elfwing, S., Seymour, B., 2017. Parallel reward and punishment control in humans and robots: Safe reinforcement learning using the MaxPain algorithm, in: 2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), IEEE, Lisbon. pp. 140–147. URL: http://ieeexplore.ieee.org/document/8329799/, doi:10.1109/DEVLRN.2017.8329799.

[5] Eoh, G., Park, T.H., 2021. Cooperative Object Transportation Using Curriculum-Based Deep Reinforcement Learning. Sensors 21, 4780. URL: https://www.mdpi.com/1424-8220/21/14/4780, doi:10.3390/s21144780.

[6] Hart, P., Nilsson, N., Raphael, B., 1968. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Transactions on Systems Science and Cybernetics 4, 100–107. URL: http://ieeexplore.ieee.org/document/4082128/, doi:10.1109/TSSC.1968.300136.

[7] Heindl, C., Stübl, G., Pönitz, T., Pichler, A., Scharinger, J., 2019. Visual large-scale industrial interaction processing, in: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, Association for Computing Machinery, New York, NY, USA. p. 280–283. URL: https://doi.org/10.1145/3341162.3343769, doi:10.1145/3341162.3343769.

[8] Iovino, M., Scukins, E., Styrud, J., Ögren, P., Smith, C., 2022. A survey of behavior trees in robotics and ai. Robotics and Autonomous Systems 154, 104096. URL: https://www.sciencedirect.com/science/article/pii/S0921889022000513, doi:https://doi.org/10.1016/j.robot.2022.104096.

[9] Jiang, Y., Bharadwaj, S., Wu, B., Shah, R., Topcu, U., Stone, P., 2021. Temporal-Logic-Based Reward Shaping for Continuing Reinforcement Learning Tasks. Proceedings of the AAAI Conference on Artificial Intelligence 35, 7995–8003. URL: https://ojs.aaai.org/index.php/AAAI/article/view/16975, doi:10.1609/aaai.v35i9.16975.

[10] Lamon, E., Fusaro, F., Balatti, P., Kim, W., Ajoudani, A., . A visuo-haptic guidance interface for mobile collaborative robotic assistant (moca), in: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE. pp. 11253–11260.

[11] Liang, E., Liaw, R., Nishihara, R., Moritz, P., Fox, R., Goldberg, K., Gonzalez, J., Jordan, M., Stoica, I., 2018. RLlib: Abstractions for Distributed Reinforcement Learning, in: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, PMLR. pp. 3053–3062. URL: https://proceedings.mlr.press/v80/liang18b.html.

[12] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal Policy Optimization Algorithms URL: https://arxiv.org/abs/1707.06347, doi:10.48550/ARXIV.1707.06347. publisher: arXiv Version Number: 2.

[13] Sirintuna, D., Giammarino, A., Ajoudani, A., 2022. Human-Robot Collaborative Carrying of Objects with Unknown Deformation Characteristics, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, Kyoto, Japan. pp. 10681–10687. URL: https://ieeexplore.ieee.org/document/9981948/, doi:10.1109/IROS47612.2022.9981948.

[14] Sirintuna, D., Giammarino, A., Ajoudani, A., 2023a. An object deformation-agnostic framework for human–robot collaborative transportation. IEEE Transactions on Automation Science and Engineering , 1–14doi:10.1109/TASE.2023.3259162.

[15] Sirintuna, D., Ozdamar, I., Gandarias, J.M., Ajoudani, A., 2023b. Enhancing human-robot collaboration transportation through obstacle-aware vibrotactile feedback.

[16] Sonar, A., Pacelli, V., Majumdar, A., 2020. Invariant Policy Optimization: Towards Stronger Generalization in Reinforcement Learning URL: https://arxiv.org/abs/2006.01096, doi:10.48550/ARXIV.2006.01096. publisher: arXiv Version Number: 3.

[17] Song, Y., Li, Y.b., Li, C.h., Zhang, G.f., 2012. An efficient initialization approach of Q-learning for mobile robots. International Journal

of Control, Automation and Systems 10, 166–172. URL: http://link.springer.com/10.1007/s12555-012-0119-9, doi:10.1007/s12555-012-0119-9.

[18] SunWoo, Y., Lee, W., 2021. Comparison of deep reinforcement learning algorithms: Path Search in Grid World, in: 2021 International Conference on Electronics, Information, and Communication (ICEIC), IEEE, Jeju, Korea (South). pp. 1–3. URL: https://ieeexplore.ieee.org/document/9369800/, doi:10.1109/ICEIC51217.2021.9369800.

[19] Sutton, R.S., 1990. Integrated Architectures for Learning, Planning, and Reacting Based on Approximating Dynamic Programming, in: Machine Learning Proceedings 1990. Elsevier, pp. 216–224. URL: https://linkinghub.elsevier.com/retrieve/pii/B9781558601413500304, doi:10.1016/B978-1-55860-141-3.50030-4.

[20] Tuci, E., Alkilabi, M.H.M., Akanyeti, O., 2018. Cooperative Object Transport in Multi-Robot Systems: A Review of the State-of-the-Art. Frontiers in Robotics and AI 5, 59. URL: https://www.frontiersin.org/article/10.3389/frobt.2018.00059/full, doi:10.3389/frobt.2018.00059.

[21] Xiao, Y., Hoffman, J., Amato, C., 2021. Macro-Action-Based Deep Multi-Agent Reinforcement Learning. URL: http://arxiv.org/abs/2004.08646. arXiv:2004.08646 [cs].

[22] Yu, X., He, W., Li, Y., Xue, C., Li, J., Zou, J., Yang, C., 2021. Bayesian Estimation of Human Impedance and Motion Intention for Human–Robot Collaboration. IEEE Transactions on Cybernetics 51, 1822–1834. URL: https://ieeexplore.ieee.org/document/8879539/, doi:10.1109/TCYB.2019.2940276.