Supplementary Material: Difficulty-Aware Time-Bounded Planning under Uncertainty for Large-Scale Robot Missions

Michal Staniaszek, Lara Brudermüller, Raunak Bhattacharyya, Bruno Lacerda, Nick Hawes

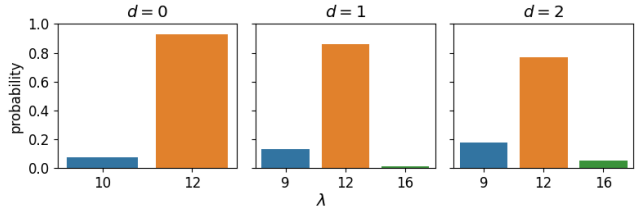

Fig. 1. Categorical distributions for durations $dur_{serv}(v, 0, 1, d)$ for each difficulty level. Recall that *difficulty* here corresponds to high, medium, and low *confidence* in the robot's localisation uncertainty. Lower difficulty has higher duration values because we are not using the optimal distance from the target surface. Higher difficulty levels imply higher localisation error which sometimes results in the robot being closer to the surface, so it can apply the dose more effectively. This is because we do not use the optimal irradiation distance, as explained in Sec. I-B.2. Note that in this setting, the duration distributions are actually independent of the location v.

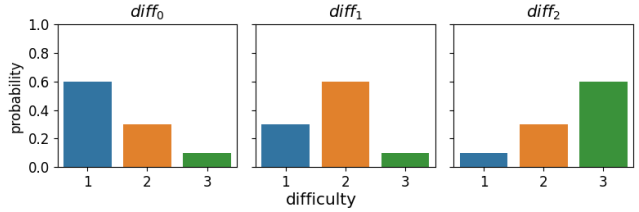

Fig. 2. Possible distributions for $diff(v)$ in the cleaning domain, indicating how dirty a room is likely to be. One of these distributions is randomly assigned to each location in the map.

I. UV Domain

For the case study in the UV disinfection domain we assume the robot hardware in Fig. 3, where the robot carries a panel of UVC light strips, with each strip providing an array of point light sources. The target environment is the library shown in Fig. 1 in the original paper, which is translated into a manually constructed topological map.

A. Preliminaries on UV Disinfection

The total UV dosage received at a point on a surface is defined by $D = I \cdot \Delta t$ (expressed in J/m^2 $= (\mathrm{W} \cdot s)/\mathrm{m}^2$), where I measures the intensity of the UVC light the point receives and Δt corresponds to the time duration of exposure. The intensity at a point is proportional to the inverse square distance to the robot [1]. When treating the source of radiation as a point light source, I can be expressed as Eqn. 1, where P_{UVC} corresponds to the power rating of the UVC light, $\eta \leq 1$ defines an attenuation factor and r represents the distance from the light source to the sample point. For our model, we use a conservative estimate of $\eta \approx 0.1$; and P_{UVC} is 8W per light source according to our hardware specifications.

$$I = \frac{P_{UVC}}{A_{exposed}} = \frac{\eta P}{4\pi r^2} \tag{1}$$

B. UV Modeling Details

1) Service levels: A 1-log reduction corresponds to 90% inactivation and D_{90} represents the dosage needed to achieve this log-reduction. Since UVC disinfection is log-linear with respect to time, reaching a 99% inactivation rate takes twice the exposure time, i.e. $D_{99} = 2 \cdot D_{90}$. Recently, multiple studies have attempted to determine dosage thresholds for inactivation of SARS-CoV-2, but values are not yet consistent across studies as they are collected across different sources, and do not follow the log-linear property [1]–[4]. In our model, we use the 1-log reduction threshold of [1] to compute values for further reductions. We have service level 1 as $D_{90} = 100\mathrm{J/m}^2$ with reward R_1, service level 2 as $D_{99} = 200\mathrm{J/m}^2$ with reward $R_1/2$, and service level 3 as $D_{99.9} = 400\mathrm{J/m}^2$ with reward $R_1/4$, where $R_1 = 100$.

2) Irradiation Model: We next define an irradiation model to relate the time the robot spends in one location to the UV dose delivered to the associated surface.

We assume that at each location in the topological map the robot only has to disinfect one surface, a shelf it is parked in front of, and approximate the surface of a shelving unit as a plane in 3D space. Using the physical model above, we derive an irradiation model that can deal with any plane in 3D space. We calculate the dose received on this plane by discretising it into a grid of surface patches. Per grid cell of the shelf surface grid, we (i) iterate over all UV point light sources of the robot, (ii) compute whether the cell's centre point is in the field of view of the source (corresponding to a 45 degree cone), and if so (iii) calculate the dosage D received from that source. The dosage value associated to a grid cell is the sum of dosage values received from all point light sources. The localisation uncertainty of the robot affects its pose relative to the shelf. We rotate and translate the shelf relative to the robot, as shown in Fig. 3. We use the first quartile of all dosage values of all grid cells to quantify dosages received on a shelf surface using a single value. This gives a conservative estimate of the entire dosage. Fig. 4 shows a heatmap of the normalised dosage values of the discretised shelf surface when the robot radiates at a distance of 1.5m without any rotation or translation from the target pose. Analysing the effects of varying the robot pose in x, y and θ in more detail, we find

All authors are with the Oxford Robotics Institute, University of Oxford, Oxford, UK {michal, larab, raunakbh, bruno, nickh}@robots.ox.ac.uk

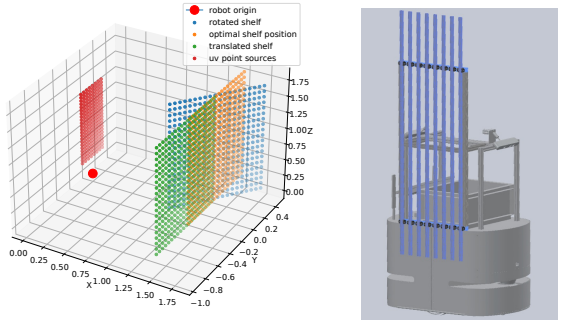

Fig. 3. *Left:* Outline of the robot setup and how uncertainties in robot position and orientation are projected onto the shelf pose. *Right:* CAD model of the robot setup chosen after considering the given radiation model.

that nodes in the topological map should ideally be placed at locations with a distance $d = 0.31$m from each shelf in order to achieve the highest possible radiation performance. However, this is hard to achieve in practice. In addition, placing nodes at this distance resulted in a single value in the categorical distribution of $P(dur_{serv}(v, l, l', d))$. As such, to make the physical scenario more plausible given robot motion behaviour and to ensure a wider distribution of service durations, we assume that nodes are $1m$ from the target surface.

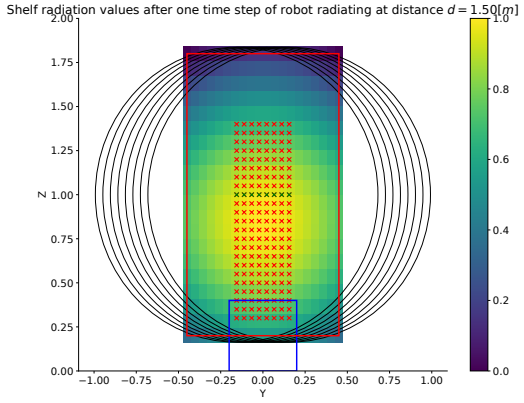

Fig. 4. Example of normalized radiation values for each grid cell of a discretised shelf surface (red rectangle) radiated by the robot (blue) at a distance of 1.5 meters. The black circles correspond to the fields of view at the intersection of the shelf plane for the green array of point sources (crosses).

3) Transition functions: Below we describe how the probabilistic evolution of state is encoded in the transition functions for the difficulty and duration state factors *diff* and dur_{serv}.

Transition function for diff: Following the *Augmented MDP* approach [5], [6], we augment the state representation by adding a variable for the *difficulty* represented by the robot's localisation uncertainty. To learn the localisation dynamics of the robot, we recorded a dataset from a deployment of a *MetraLabs SCITOS X3* robot[1] navigating on the given topological map of the library over a period of approximately 4 hours. The robot localised using Adaptive Monte Carlo Localisation (AMCL) [7] and the dataset includes logs of the robot's localisation uncertainty from this method. We construct the difficulty levels from a dataset $\{(\sigma_{x_0}, \sigma_{y_0}, \sigma_{\theta_0}), \ldots, (\sigma_{x_m}, \sigma_{y_m}, \sigma_{\theta_m})\}$ containing m samples of pose standard deviations. We cluster the standard deviations using a Gaussian Mixture Model (GMM), defined as $P(\vec{\sigma}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\vec{\sigma}|\vec{\mu}_i, \Sigma_i)$, where K sets the number of multivariate Gaussian components and $\vec{\phi}$ is a vector of weights for each component in the mixture model. Each component is a separate multivariate Gaussian distribution modelling variations in x, y, θ. We map each component to a difficulty level in the MDP, in order of increasing norm of the component's mean, so the lowest difficulty is the one with lowest localisation variance. The number of components in the GMM (and hence the number of difficulty levels D) was determined using the Bayesian Information Criterion. We assume that the difficulty at a location is independent of all other state factors. To compute $P(diff(v))$ for a location, we assign all pose standard deviations sampled at location v to a component of the GMM according to maximum likelihood. The probability of each difficulty d for the location is then the proportion of samples associated with that difficulty's GMM component.

Algorithm 1 Generation of $P(dur_{serv}(v, l = 0, l' = 1, d))$

Given: RadiationModel
Parameters: `num_confidence_levels`, `num_samples`, user-defined dwell pose $\boldsymbol{x}^*$
Init: $all\_dists \leftarrow \emptyset$
Fit GMM(`num_confidence_levels`) to stdevs (AMCL data)
Order means of GMM components by their norm along each dimension
for $\sigma_c = (\mu_x, \mu_y, \mu_\theta)$ in ordered GMM components **do**
  $durations_c \leftarrow \emptyset$
  $i \leftarrow 0$
  while $i \leq$ `num_samples` **do**
    Sample displacement: $\Delta x \sim \mathcal{N}(0, \sigma_c)$
    $\boldsymbol{x'} = \boldsymbol{x}^* + \Delta x$
    $d_{\Delta t=1}$ = RadiationModel($\boldsymbol{x'}$)
    $dur^i_{serv}(v, 0, 1, c) = \texttt{clean}/d_{\Delta t=1}$
    $durations_c \leftarrow durations_c \cup \{dur^i_{serv}(v, 0, 1, c)\}$
    $i \leftarrow i + 1$
  end while
  $all\_dists \leftarrow all\_dists \cup durations_c$
end for
return all_dists

Transition function for dur_{serv}: The service duration is the duration needed to reach cleanliness level l' from l. The probability of the duration being observed is modelled by $P(dur_{serv}(v, l, l', d) = \lambda)$, i.e. the duration required to clean to a given level is defined by the level l' chosen and the

[1] `https://www.metralabs.com/en/mobile-robot-scitos-x3/`

difficulty d, in this case the robot's localisation uncertainty. For each d, we build $dur_{serv}(v, 0, 1, d)$ by sampling poses from the GMM component for d and using the UV irradiation model described above to determine the duration required to bring the lower dose quartile to the service level 1. We then fit a categorical distribution to these samples using a discretisation of 5s. This gives us distributions for applying a dose of $100\mathrm{J/m}^2$ to a location. Since the service levels are log-linear, $dur_{serv}(v, 0, 2, d) = 2 \cdot dur_{serv}(v, 0, 1, d)$, and $dur_{serv}(v, 0, 3, d) = 4 \cdot dur_{serv}(v, 0, 1, d)$, so they do not need to be separately computed. These distributions are used in the simplified versions with the single service actions.

This overall procedure is described in Alg. 1, and an example for the resulting categorical distributions is shown in Fig. 1.

II. Experimental Evaluation

A. Policy Translation

To evaluate policy quality, we use a discrete event simulator. Execution of a policy in the simulator requires a mapping of states and actions from the model which generated the policy (the *policy model*) to the MDP which is representing the world (the *world model*). For example, the world model I-AD-FN has state space $s = (v_i, l_1, \ldots, l_n, d_i, \tau)$ and policy model S-ST-TSP has $s = (v_i, l, d, \tau)$. To translate from policy to world state, we copy all elements, ensuring that l in the policy state is mapped to l_i in the world state. From policy to world, we again copy all elements, mapping only l_i to l. The single service action of the policy model must be translated into up to 3 incremental service actions in the world model. $a_{0,2}$ on the policy would become the action set $\{a_{0,1}, a_{1,2}\}$ on the world.

Edge actions in the policy model need to be translated to execute the full route between two nodes as individual edge actions, rather than the consolidated shortest path action used in the fixed navigation model. With $e = (0, 3)$, a_e might translate to several edge actions with edges $\{(0, 1), (1, 2), (2, 3)\}$.

For the translation to function, it is critical that applying the same service action $a_{l_i,l'}$ in a given state in the world and policy models results in an identical set $\lambda_{a_{l_i,l'}} \subseteq \Lambda$ of possible outcomes for the change in elapsed time τ. If this is not the case, there will be a mismatch between valid states in the world model and policy model, and the policy being executed will not contain an action for a state produced by the world model.

B. Map connectivity may affect rewards

Table I shows the results of 1000 policy executions on the partially and fully connected 6 location map. For the shortest time horizon, being fully connected allows a slightly higher reward to be gathered. The difference between the two distributions of rewards is significant in the case of I-ST-FN ($p = 5.7 \times 10^{-27}$) and S-ST-FN ($p = 3.1 \times 10^{-10}$) according to a two-sided Kolmogorov-Smirnov test. We hypothesise that the difference between free and fixed navigation would be more obvious on larger maps, but are unable to test this due to memory constraints.

TABLE I

Total rewards for different time bounds on the tiny map in the UV domain, across 1000 policy executions, P indicates tiny map, F indicates tiny_fc map

Model	Conn	$\alpha_1 = 1$	$\alpha_1 = \alpha_2 = 0.5$	$\alpha_2 = \alpha_3 = 0.5$
Baseline	P	68 ± 79	315 ± 61	424 ± 32
	F	69 ± 75	317 ± 62	424 ± 32
I-AD-FN	P	516 ± 29	722 ± 27	962 ± 13
	F	522 ± 29	722 ± 27	N/A
I-ST-FN	P	506 ± 23	711 ± 23	956 ± 11
	F	518 ± 28	712 ± 23	956 ± 11
S-ST-FN	P	505 ± 22	705 ± 23	952 ± 12
	F	513 ± 25	703 ± 25	953 ± 12
I-AD-TSP	P	513 ± 31	720 ± 25	960 ± 13
	F	513 ± 31	719 ± 25	961 ± 13
I-ST-TSP	P	506 ± 22	708 ± 23	954 ± 10
	F	505 ± 20	707 ± 22	955 ± 11
S-ST-TSP	P	506 ± 22	703 ± 24	952 ± 12
	F	505 ± 20	702 ± 24	952 ± 12

ACKNOWLEDGMENT

This work was supported by the EPSRC Programme Grant "From Sensing to Collaboration" (EP/V000748/1), the UKAEA/EPSRC Fusion Grant (EP/W006839/1), and a gift from Amazon Web Services.

References

[1] A. Pierson, J. W. Romanishin, H. Hansen, L. Z. Ya, and D. Rus, "Designing and Deploying a Mobile UVC Disinfection Robot," in *IROS*, 2021, p. 2.

[2] H. Kariwa, N. Fujii, and I. Takashima, "Inactivation of SARS coronavirus by means of povidone-iodine, physical conditions and chemical reagents," in *Dermatology*, vol. 212, 2006, pp. 119–123.

[3] J. M. Correia Marques, R. Ramalingam, Z. Pan, and K. Hauser, "Optimized Coverage Planning for UV Surface Disinfection," in *ICRA*, 2021.

[4] M. Biasin, A. Bianco, G. Pareschi, A. Cavalleri, C. Cavatorta, C. Fenizia, P. Galli, L. Lessio, M. Lualdi, E. Tombetti *et al.*, "Uv-c irradiation is highly effective in inactivating sars-cov-2 replication," *Scientific Reports*, 2021.

[5] L. Nardi and C. Stachniss, "Uncertainty-aware path planning for navigation on road networks using augmented MDPs," in *ICRA*, 2019.

[6] N. Roy and S. Thrun, "Coastal navigation with mobile robots," in *NeurIPS*, 2000.

[7] D. Fox, "Kld-sampling: Adaptive particle filters," *NeurIPS*, vol. 14, 2001.