Using Spatio-Temporal Affordances to Represent Robot Action Semantics

Francesco Riccio[1]* Roberto Capobianco[1]* Daniele Nardi[1]

***Abstract*— In this paper we rely upon the concept of Spatio-Temporal Affordances (STA) to formalize the objective function to learn affordance descriptors. Such a function allows to better encode action semantics related to the environment. We qualitatively evaluate obtained results over the learned spatial model for two different tasks.**

I. Introduction

Affordances have been introduced as action opportunities that objects offer [1] and, recently, they have been used in robotics to improve object representations in relation to actions. Although the original definition is limited to objects, this concept has been extended to describe environments as a combination of spatial affordances [2], [3]. In Kapadia et al. [4], for example, these are used for collision avoidance, while Diego et al. [5] use affordances for robot navigation in crowded environments.

Typically, approaches adopted in literature are rather specialized and cannot be used for the execution of multiple tasks. To tackle this problem, we use the concept of spatial semantics to establish a connection between the environment and spatio-temporal affordances. In particular, we formalize a Spatio-Temporal Affordance Map (STAM) [6] as a representation that contains high-level semantic properties of the operational scenario. These properties are encoded within a set of descriptors that can provide prior information about actions and that can be easily learned.

In this paper, we describe and formalize the objective function to be optimized in order to learn and obtain affordance descriptors. By using this approach, such descriptors can be easily learned and integrated within a behavior learning framework. In particular, in our experiments we learned the appropriate relative distances of a robot, as a function of the state of the environment, for a following and a reaching task on a NAO robot. The former was learned by observing expert policies – from demonstration, while the latter was incrementally learned during the iterations of a Monte-Carlo reinforcement learning method. As shown in Fig. 1, the obtained parameters accurately model action semantics as a function of the state of the environment and can be used to guide an autonomous STAM agent during the execution of tasks.

[1]Francesco Riccio, Roberto Capobianco and Daniele Nardi are with the Department of Computer, Control, and Management Engineering, Sapienza University of Rome, via Ariosto 25, Rome, 00185, Italy {riccio,capobianco,nardi}@dis.uniroma1.it

* These two authors equally contributed to the work.

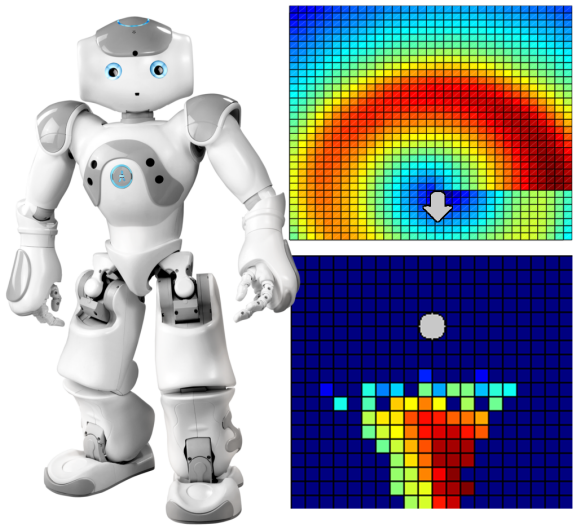

Fig. 1. Heatmaps representing the spatio-temporal affordance models for a "following" (top) and a "reaching" (bottom) task on a NAO robot. Areas of the environment that better afford the tasks are highlighted with red colors. In the top image, the arrow represents the target to be followed and points in its walking direction. The learned spatial affordance spreads around the target defining the positions for the robot that supports the execution of the task. In the bottom image, instead, the circle represent the target to be reached and the spatial distribution highlights the areas of the environment that afford a "move forward" action.

II. STAM: Spatio-Temporal Affordance Map

Manifold works demonstrate that a proper spatial semantic representation can improve robot capabilities. Typically, approaches in literature [2]–[5] exploit spatial semantics to support the execution of a single specific task. For example, Rogers et al. [8] and Kunze et al. [9] exploit semantic knowledge to afford a search task. Instead, Epstein et al. [2] employ spatial affordances to support navigation, while Luber et al. [3] use them to improve the performance of a tracking system. Conversely, we rely upon a general architecture that enables to simultaneously model the spatial affordances of different tasks through a modular approach. To this end, we introduce the concept of Spatio-Temporal Affordance (STA) as a general function that defines areas of the operational environment affording a task, given a state of the world.

Definition 1: A *Spatio-Temporal Affordance (STA)* is a function

$$f_{E,\mathcal{T}} : S \times \Theta \rightarrow \mathrm{A}_E. \tag{1}$$

$f_{E,\mathcal{T}}$ depends on the environment E and a set of available tasks $\mathcal{T} = \{\tau(t)\}$. It takes as input the state of the

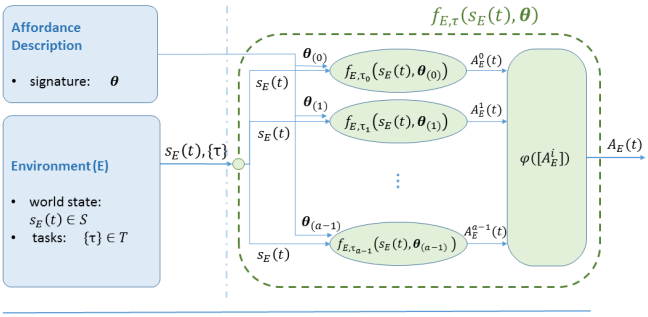

Fig. 2. Decomposition and detailed overview of a Spatio-Temporal Affordance Map (STAM).

environment $s_E(t) \in S$ at time t, a set of parameters $\boldsymbol{\theta} \in \Theta$ characterizing the affordance function, and outputs a map of the environment A_E that evaluates the likelihood of each area of E to afford $\mathcal{T}$ in s_E at time t.

Such function constitutes the core element of a Spatio-Temporal Affordance Map (STAM). It takes as input both the parameters $\boldsymbol{\theta}$ obtained from an affordance description module (*a-module*) and the state of the world from the environment module (*e-module*), which also defines the possible tasks $\mathcal{T}$. In particular:

- the a-module is a library of parameters $\boldsymbol{\theta}$ representing the *signature* of the STA. Such signature defines the spatial distribution of affordances within the environment;
- the e-module encodes the state of the world $s_E(t)$ provided to the STA function, as well as the set of available tasks $\mathcal{T}$.

As shown in Fig. 2, STAM can be used both to interpret relations among different affordances (if any) and to represent affordances individually. In fact, a STA can be seen as a composition of different $f_{E,\tau_j}(s_E(t), \boldsymbol{\theta}(j))$ functions, with $j \in [0, a-1]$, where a is the number of affordances, each modeling the spatial distribution A_E^j of a particular affordance in E. These are then combined by a function ϕ, that takes as input all the A_E^j and outputs a map A_E supporting the execution of $\mathcal{T}$.

A. Learning the STA Signature

According to Def. 1, the generation of A_E depends on the signature $\boldsymbol{\theta}$. In fact, it directly modifies the way the space is modeled and constitutes the main vehicle to shape affordances. For example, we employ a Gaussian Mixture Model (GMM) to implement the function $f_{E,\mathcal{T}}$, and we represent the signature $\boldsymbol{\theta}$ of the STA function as a set $\boldsymbol{\theta} = \langle \pi_1, \mu_1, \Sigma_1, \ldots, \pi_N, \mu_N, \Sigma_N \rangle$, where π_i is the prior, μ_i the mean vector and Σ_i the covariance matrix of a mixture of N Gaussians. Hand-designing such signature is difficult, and requires an accurate understanding of both the parameters $\boldsymbol{\theta}$ and the function $f_{E,\mathcal{T}}$. Conversely, it can be learned by observing a dataset $\mathcal{D}$ of state-task pairs and by exploiting expert demonstrations or a general reinforcement learning paradigm. As shown in Eq. 2, such signature is chosen to maximize, for each task $\tau \in \mathcal{T}$, the summed likelihoods to afford τ in all the states labeled with τ in $\mathcal{D}$:

$$\boldsymbol{\theta}(j)_i = \arg\max_{\boldsymbol{\theta}(j)} \sum_{\{s|(s,\tau_j)\in\mathcal{D}\}} f_{E,\tau_j}(s_t, \boldsymbol{\theta}(j)), \qquad (2)$$

where $j = 1 \ldots |\mathcal{T}|$.

Intuitively, the higher the likelihood to afford the tasks in $\mathcal{D}$, the better the function signature encodes the spatial semantics of the robot tasks in E.

III. Discussion

Spatio-temporal affordances allow to encode robot action semantics directly into the operational scenario. Fig. 1 reports a qualitative evaluation of a "following" and "reaching" tasks formalized though the STAM architecture and the function signature here proposed. In the "following" case, for example, the spatial distribution of the affordances defines the area around the target that better support the task to follow a person. In the latter case, equivalently, the affordances define the area in front of the target as the most promising one in order to reach the desired object. It is worth remarking that, in both cases, by maximizing the affordance function as in Eq. 2, we successfully exploit spatial affordances to enable a STAM agent to determine the best position in E according to the task to perform and a given state of the environment.

References

[1] J. J. Gibson, *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1979.

[2] S. L. Epstein, A. Aroor, M. Evanusa, E. Sklar, and S. Parsons, "Navigation with learned spatial affordances," in *COGSCI*, 2015.

[3] M. Luber, G. D. Tipaldi, and K. O. Arras, "Place-dependent people tracking," *Int. Journal of Robotics Research*, vol. 30, no. 3, pp. 280–293, 2011.

[4] M. Kapadia, S. Singh, W. Hewlett, and P. Faloutsos, "Egocentric affordance fields in pedestrian steering," in *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games*, ser. I3D '09. New York, NY, USA: ACM, 2009, pp. 215–223.

[5] G. Diego and T. K. O. Arras, "Please do not disturb! minimum interference coverage for social robots," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 1968–1973.

[6] F. Riccio, R. Capobianco, M. Hanheide, and D. Nardi, "Stam: A framework for spatio-temporal affordance maps," in *Proceedings of the 2016 Modelling and Simulation for Autonomous Systems (MESAS'16) Workshop*, 2016.

[7] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory–motor coordination to imitation," *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, Feb 2008.

[8] J. G. Rogers and H. I. Christensen, "Robot planning with a semantic map," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, May 2013, pp. 2239–2244.

[9] L. Kunze, C. Burbridge, and N. Hawes, "Bootstrapping probabilistic models of qualitative spatial relations for active visual object search," in *AAAI Spring Symposium 2014 on Qualitative Representations for Robots*, Stanford University in Palo Alto, California, US, March, 24–26 2014.

[10] D. V. Lu, D. Hershberger, and W. D. Smart, "Layered costmaps for context-sensitive navigation," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2014, pp. 709–715.

[11] A. Chemero, "An outline of a theory of affordances," *Ecological Psychology*, vol. 15, no. 2, pp. 181–195, 2003.

[12] S. Gelly and D. Silver, "Combining online and offline knowledge in uct," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 273–280.