Active Learning in Robotics: A Review of Control Principles

Annalisa T. Taylor, Thomas A. Berrueta, and Todd D. Murphey[1,]

[a]*Mechanical Engineering, Northwestern University, 2145 Sheridan Rd., Evanston, 60208, Illinois, United States*

Abstract

Active learning is a decision-making process. In both abstract and physical settings, active learning demands both analysis and action. This is a review of active learning in robotics, focusing on methods amenable to the demands of embodied learning systems. Robots must be able to learn efficiently and flexibly through continuous online deployment. This poses a distinct set of control-oriented challenges—one must choose suitable measures as objectives, synthesize real-time control, and produce analyses that guarantee performance and safety with limited knowledge of the environment or robot itself. In this work, we survey the fundamental components of robotic active learning systems. We discuss classes of learning tasks that robots typically encounter, measures with which they gauge the information content of observations, and algorithms for generating action plans. Moreover, we provide a variety of examples—from environmental mapping to nonparametric shape estimation—that highlight the qualitative differences between learning tasks, information measures, and control techniques. We conclude with a discussion of control-oriented open challenges, including safety-constrained learning and distributed learning.

Keywords: Active learning, Robotics, Robot control, Learning theory, Perception and sensing, Artificial intelligence

1. Introduction

"Perceptual activity is exploratory, probing, searching; percepts do not simply fall onto sensors as rain falls onto ground. We do not just see, we look." (R. Bajcsy in her 1988 paper *Active Perception* [1]). The difference between seeing and looking is the presence of action—seeing is passive and looking is active. Unfortunately, we do not use distinct words for passive learning and active learning, often leading to confusing the two and unintentionally treating "learning" as passive learning with active learning as an afterthought. Nevertheless, how we acquire data impacts the quality of learning and what is even possible to learn, indicating that control—both analysis and synthesis—in learning will inevitably be important. More than three decades after Bajcsy's comments, the key elements of how control synthesis and analysis should inform learning remain largely unaddressed, and the vast majority of work in learning still focuses on analysis of passively collected data; this body of work makes up a statistical theory of learning. Still absent is an action-oriented theory of learning—a control theory for learning. How should control synthesis affect learning? What sort of feedback interconnections facilitate learning?

When prior knowledge and existing datasets are widely available, passive learning has proven to be a successful tool for constructing parametric representations of statistical relationships in data. Broadly, passive learning is an optimization process in which the parameters of a model are fit according to data. The last decade has seen major strides in robotics dependent on the advent of modern learning methodologies, particularly variations of deep neural networks [2]. However, in settings where previously existing data sets are unavailable, and where products of human knowledge (*e.g.*, labeled datasets, knowledge graphs) do not exist, a robot will have to engage in unsupervised discovery and acquire the data it needs [3]. We refer to this process as *active* learning (see Figure 1). In contrast to passive learning, active learning is a decision-making process where agents take actions to gather the data that best realizes a learning objective.

Animals use their bodies to learn. To paraphrase Bajcsy, we do not just passively learn, we actively learn—the pages of a book do not just turn before our eyes while we absorb information. For agents with physical bodies, such as animals or robots, active learning demands understanding and exploiting the role of embodiment and physical interaction in learning. Insofar as robotics should take inspiration from biology, active learning in robotics will involve the purposeful movement of a robot's body; here, control synthesis tools will connect decision-making to the resulting movement.

There is a rich literature on how animals use their bod-

Email address: `annalisa.taylor@u.northwestern.edu`, `tberrueta@u.northwestern.edu`, `t-murphey@northwestern.edu` (Annalisa T. Taylor, Thomas A. Berrueta, and Todd D. Murphey)

URL: `murpheylab.github.io` (Annalisa T. Taylor, Thomas A. Berrueta, and Todd D. Murphey)

[1] This material is based upon work supported by the United States National Science Foundation under Grant CNS 1837515 and by the United States Army Research Office MURI award W911NF-19-1-0233. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the aforementioned institutions.

ies and movements to improve information acquisition [4–12]. For example, in [13] we demonstrated that a variety of animals engage in active information acquisition by exploring their environment in proportion to the local amount of perceived uncertainty. In addition to a medium for embodied movement plans, physical bodies are independently capable of implicit computation [14, 15], information storage [16], novelty detection [17], and learning [18]. By harnessing the power of embodiment and morphological computation [19], active learning presents a promising way forward for robotics problems where the outcomes of physical interactions may be unknown *a priori*, such as in soft robotics [20].

Not only is embodiment and movement paramount to information acquisition and active learning, but movements themselves can be informative. Recent work analyzing animal and human movement has begun to interpret physical bodies as information channels and motions as information-carrying signals. This has led to the development of methods that help to understand the pathology of conditions such as autism spectrum disorder [21], schizophrenia [22], and stroke [23, 24] through an information-theoretic analysis of movement. More generally, this suggests that in order to realize learning objectives, active learning requires measures that capture the information content of an agent's movements.

Counterintuitively, information-rich movement does not always appear productive, orderly, or carefully planned. A well-studied example of this is the optimality of diffusion in animal foraging—here, purely stochastic motion plans have been shown to be highly informative [25–27]. Another example of interest to researchers for decades is that of playful behavior in animals [28, 29]. One may ask why animals would expend significant energy on movement that is not key to survival; for our purposes, we consider these active behaviors as enhancing learning [30]. Hence, to learn through movement, agents must engage in exploratory behaviors that may not always seem useful.

Despite its clear connections to our understanding of learning in animals and humans, the field of active learning finds its origins in theoretical computer science [31]. In this setting, agents are represented by disembodied algorithms whose actions are limited to making queries about observed data samples. As a result, many modern frameworks for artificial intelligence have tended to neglect the role of physics and embodiment on the learning process. However, adapting to the constraints of the real world is crucial to learning in the wild. Even the most successful traditional machine learning techniques for robot control, such as reinforcement learning, rely on "big data" generated from simulated rollouts. In reality, robot deployment is a time and physically intensive activity, and robots cannot be instantly reset and redeployed at will. To make matters worse, informative data samples are typically sparse. Taken together, these issues highlight the importance of considering sample efficiency and deployment efficiency in robot learning. On the other hand, control theory has a long history of dealing with the constraints imposed by the laws of physics, while simultaneously managing secondary—yet very important—objectives such as safety, robustness, and efficiency.

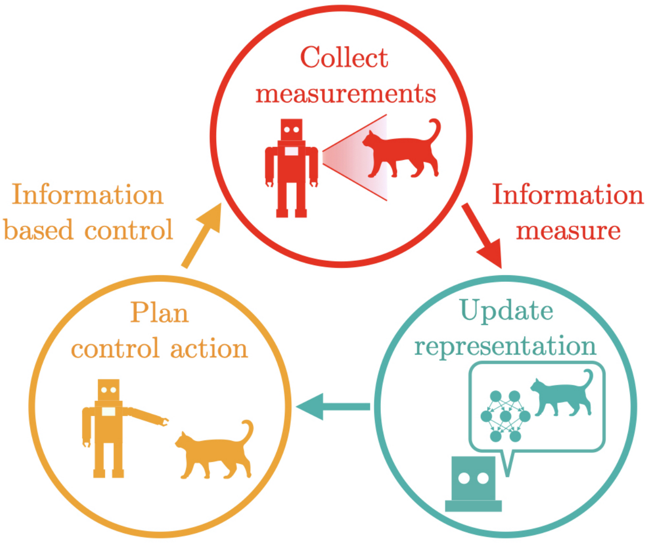

Figure 1: **The active learning process:** A learner leverages information measures to formulate actions for collecting relevant or descriptive data. Active learning includes the feedback control of a system for which the internal state is both a learning system and history.

There are many areas of robotics that will require the type of black-box flexibility of machine learning to make progress. When principled alternatives to modeling the physics of complex interactions between agents and their environments do not exist, machine learning can sometimes be the only way to enable robot control. One area in which flexible learning tools are particularly useful is in high-dimensional nonlinear sensing, where deep convolutional networks excel at integrating potentially hundreds of complex and highly-redundant sensory signals into compressed and informative signals [15, 32]. At times, the interactions between a robot and its environment may be infeasible to model either due to properties of the environment (*e.g.*, locomotion in granular media [33]), or the robot itself (*e.g.*, compliant soft robots [34]). Thus, the field of robotic active learning has the potential to overcome the challenges inherent to robot control and machine learning by inheriting the best qualities of both. In this review, we highlight important progress made towards this goal, and motivate future directions for developing an action-oriented theory of embodied learning.

The organization of this review is as follows. First, we cover the history and basic considerations required for an active learning system—what there is to learn, how to measure information in actions, and how to generate such informative actions. Then, we survey key areas of application for active learning and open challenges in the field. The authors' own work plays a role in creating a narrative, but with consistent reference to the broad literature in robotics on related areas. Section 2 covers a brief history of the field of active learning and its origins in the broader field of computer science. In Section 3 we cover different learning goals, increasing in complexity from learning state parameters to abstract features. Then, Section 4 discusses measures of information, focusing on those appropriate to be used as control objectives by synthesis methods such as those in Section 5. In Section 6, we discuss common applications where facets of active learning naturally arise, whether explicitly or implicitly, in problem formulations. Finally, in Section 7 we discuss extensions and open challenges followed by conclusions in Section 8.

2. History Of Active Learning

Since its inception, robotics has been interested in making embodied agents learn and adapt to their surroundings like biological organisms [35]. However, due to fundamental limitations on computing hardware, programming machines [36] and adaptation to external stimuli [37], robot learning was limited to the most rudimentary demonstrations throughout the mid-20th century. After establishing his theory of computation [38], Alan Turing shifted his focus to the question of whether machines could think and learn [39]. Turing's efforts prompted both the philosophical and formal study of artificial intelligence [40].

While hardware posed constraints on applied learning, the second half of the century saw the founding of the field of computational learning theory [41–44]. Analogous to computability theory, computational learning theory focuses on assessing the "learnability" of concepts under different models of learning, such as inductive inference [45], online learning [46], statistical query learning [47, 48], among many others. The diversity of models of learning speaks to the difficulty of capturing what we mean when we say that a concept is learnable. To this day, useful models of learning are being introduced to tackle new problems on learnability [49]. Of the many mathematical frameworks for learning, the most successful and widely used is the Probably Approximately Correct (PAC) learning model [43, 50, 51]—a particularly important framework because it was the first to bring insights from the theory of computational complexity to the study of learning. Across its many models of learning, computational learning theory forms the primary means through which we mathematically model and formally understand learning as a computational problem.

The influence of computability theory [52] is particularly visible in the field's focus on automata theory and linguistics [53], where problems are often framed as learning languages or equivalent automata specifying the languages. In contrast, much of robotics is grounded in the history of industrial automation, where mechanical interactions are the fundamental object of interest [54]. As a result, robot learning focuses on the role of physics on sensing, actuation, and mechanical interactions with the environment for the purpose of learning.

One of the most important areas within computational learning theory is that of query learning [47, 55, 56]. This field is concerned with identifying the classes of functions that a "learner" (*e.g.*, an algorithm) can learn by observing samples of data provided by an "oracle" (*e.g.*, a teacher or an environment) using a given model of learning. At each stage of the learning process, the learner has a "learning hypothesis" about the nature of the function class that it is learning. In the context of query learning, the learner is additionally allowed to ask the oracle for information about the samples it is observing or about its current learning hypothesis [31]. The learner then must make decisions about what queries to present to the oracle in order to advance its learning objective [57]. In this way, learning is no longer framed as a passive process. Instead, it is a decision-theoretic process through which the learner takes actions in order to further its objective—or in other words, *active* learning. By leveraging their decision-making, active learners can almost always achieve the same performance as an equivalent passive learner with exponentially fewer data samples [58]. This framing can be restrictive in a robotics context where actions have the potential to elicit information and affect the environment or learning objective. Despite forming a theory grounded in the decision-making of learning agents, computational learning theory has not concerned itself with these types of practical considerations that embodied robot learning demands.

Another theory of learning largely independent from those discussed above is reinforcement learning (RL), which finds its origins in the study of conditioning in psychology [59]. As originally envisioned, RL refers to the use of external stimuli and incentive structures to elicit desired behavior out of animals or humans [60]. In this sense, RL was established as a theory of learned behavior rather than learning in-itself. However, its mathematical underpinnings were not established until the second half of the 20th century in the work of Richard Sutton and Andrew Barto among others [61–63]. By grounding their work in the theory of dynamic programming [64] and optimal control [65], Sutton and Barto created a rich mathematical theory of learning and control based on the behavioral psychology of reinforcement [66]. Typically, an RL problem is framed as a Markov Decision Process (MDP) where an agent must take actions in order to explore their environment and learn how to maximize their reward signal [67]. When agents are making decisions and taking actions to actively gather data and learn about their objective, we consider RL to be a type of active learning. In contrast, if exploration is being handled passively through naively randomized simulated experience, we do not.

Despite its early uses for optimal control [63], RL has only recently become a primary technique for robot learning due to the many successes of deep RL in continuous control [68–70]. However, most methods developed for deep RL are ill-suited to robot learning because of their large data requirements, lack of generalizability between tasks, as well as their inability to learn incrementally and guarantee safety [71–73]. While techniques such as Maximum Entropy RL have taken steps to improve data efficiency and generalizability in robot learning settings [74–76], deep RL is still far off from seamless deployment in the real world due to its reliance on simulated experience to make progress on learning and control objectives [77–79]. Moreover, easily specifying and incorporating safety [80], stability [81], controllability [82, 83], or reachability [84] remains an open challenge. Taken together, these points highlight that—despite being a theory of active learning based on the behavior of embodied agents—RL is underdeveloped for many robotic applications in its present form.

In this section we have briefly outlined the historical development of active learning as a field. Throughout the literature and across its different subfields, we have found that although researchers have had great interest in applying active learning methods to robotics problems, there is still a need for the development of theories of active learning specifically *for* robotics. Such theories of robot learning should center the properties of the agent as an embodied control system with requirements for stability, safety, sample efficiency, and continuous deployment. To this end, much of the work that we present in this review focuses on aspects of embodiment, and suggests the possibility of developing a control-oriented theory of embodied active learning.

3. What Do Robots Need To Learn?

What does a robot need to learn from data? Learning goals can be grouped into problems of increasing sophistication and level of abstraction. Here we will distinguish between learning *parameters*, as a relatively simple starting point, learning *models*, and learning *features*. This division is by no means unique, but provides a useful taxonomy for discussing what learning goals we may have for a robotic system.

3.1. Parameters

Learning parameters is relevant in many settings. For instance, one may wish to determine the location of an object, food, or predators. In this case, the parameters of interest are spatial coordinates that localize the object. If the parameters evolve in time (*e.g.*, a mobile object) they may have dynamical properties that can be exploited or learned. If a model is known, parametric filters [85–88] may be used. When the posterior probabilities of an inference model are not expected to be approximately Gaussian, nonparametric filters, such as Bayesian filters [89, 90], histogram filters [91], or particle filters [92–94] are often used instead. Active learning can be critical to overcoming sensor limitations and identifying a wide variety of parameters. A salient setting for active learning is near-field sensing. Near-field sensing includes tactile sensing which requires mechanical contact and electrosense, where close proximity is necessary. Hence, when subject to near-field sensing constraints, robots must leverage their agency for successful parameter identification. In far-field sensing, such as cameras and radar at a distance, actions may play a more limited role in parameter identification because the sensor range automatically provides substantial information without the need for movement.

3.2. Models

Models generalize parameters, and can be models of either the robot itself, such as a model of the dynamics, or the environment, such as a topographical map. The ability of a robot to learn a model of its dynamics is important in rapidly shifting environments where first-principle models struggle to make reliable predictions. The problem of system identification is often parametric, focusing on describing the dynamics using models whose structure and number of parameters are fixed *a priori*, such as in neural networks. However, system identification may be *nonparametric* as well, as in Gaussian process regression and other kernel-based methods. Nonparametric models may be particularly useful when robots operate in unstructured or unknown environments. While parametric models have also been successfully used in this context, it is difficult to know ahead of time that a parametrized model will have the representational capacity to characterize the environment. This has led to the use of models with an increasing number of parameters—sometimes on the order of billions of parameters—to ensure that the network can capture the properties of the environment.

3.2.1. Mapping

Mapping is one form of modeling the environment that emphasizes its geometry. Mapping applications often use occupancy grids [95, 96], coverage maps [91], and Gaussian process regression to represent spatially-varying phenomena or high-dimensional belief spaces [97–102]. These techniques presume coverage—that data has been taken over a sufficiently varied area to reconstruct and represent the properties of the environment. The active learning approach instead suggests that an agent reacts to data it collects locally and then adjusts its mapping strategy. While environmental mapping in open air is not an application that necessarily demands the use of active learning methods, other types of environments may not be as straightforward. For example, underwater exploration is difficult because robots are subject to stringent constraints on sensing, actuation, and communication. Here, robots often need to operate in environments where light levels prevent long-range visual monitoring, which demands the use of active learning tools in order to construct motion plans that incrementally adapt to the robot's uncertain measurements [103–105]. In [106], the authors use control and Gaussian process regression to model, map, and actively sample the distribution of phytoplankton in the

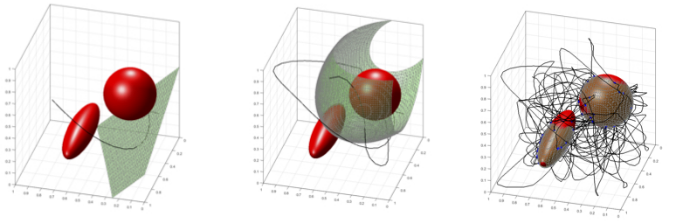

Figure 2: **Shape reconstruction:** This example shows the active identification of an unknown geometry in the environment, using binary contact measurements as the measurement modality [107]. By developing data-driven models of objects, robots can search for and recognize obstacles or tools without needing analytic or CAD models.

ocean off the coast of Norway, thereby greatly accelerating environmental monitoring and mapping of oceanic resources.

3.2.2. Shape

Similarly to mapping, nonparametric shape estimation is another area of model learning that focuses on the geometric relationship between collected data samples [108]. The shape estimation literature grew from the field of computer vision, and has traditionally focused on static tasks, such as estimating the poses of human bodies [109] or the curvature of roads from image samples [110]. However, as we increasingly deploy autonomy in the real world, determining the shape and material properties of unknown objects may be necessary to interact with them and potentially employ them as tools. To this end, the process of shape estimation may need to be dynamic and probing in nature, requiring that agents leverage their control authority to actively learn the properties of the object.

As an informative example of this kind of learning problem, we share some results from our own work. In [111], we considered nonparametric shape estimation using contact-based sensors to actively learn the shapes of obstacles in the robot's environment, which we then extended towards data-driven mapping and localization [107]. Figure 2 shows a three dimensional set of objects whose shapes are being reconstructed from binary contact measurements made by a simulated mobile robot. By actively generating trajectories that make contact with the object surfaces, we maximize the Fisher information of the support vector machine (SVM) object model and successfully identify them. The enabling insight is the use of the Fisher information, which we discuss at length in Section 4, to synthesize object-robot interactions that are optimally informative.

3.2.3. Dynamics

One of the most crucial learning tasks is that of identifying the agent's own dynamics. Whether learning the dynamics is necessary due to their intrinsic complexity, or as a result of a sudden malfunction or compliant interaction, there are many scenarios in which it may be impossible or infeasible to have an accurate prior representation of the system's dynamics. Self-identifying dynamics is an active process, where the agent needs to take actions and collect data that explore its different behavioral regimes. In some settings, models that are well-specified in certain behavioral regimes may have to be augmented through data-driven means to work in extraneous conditions. For instance, in aerospace applications data-driven techniques will have excellent data available for nominal conditions but often no data available for specific off-nominal conditions, suggesting the need for active learning outside of the nominal regime [112]. Since the literature on learning dynamics is very diverse, providing a comprehensive survey would require its own review [113–117]. Instead, here we review a few particular representations of dynamics that are of particular interest to the field of robotics.

Deep neural networks (DNNs) are models comprised of many individual units (*i.e.*, computational synthetic neurons) with limited capabilities that together, through their interconnections, are capable of great representational power [2]. As we have discussed earlier in this review, deep networks are not always suited to the demands of robot learning due to their high data and computational requirements. Nonetheless, certain network architectures have been shown to be well-suited to predicting dynamics, such as recurrent neural networks [118], whose capabilities enable them to predict the global structure of temporal dynamics from local measurements. In settings where learning does not need to occur rapidly or incrementally, carefully chosen deep learning architectures have been successful in learning robot dynamics for control [119, 120]. While DNNs have been successful in many robotic applications, the online nature of active learning tasks often prevent them from being used in these settings.

A nonparametric alternative to learning dynamics is the use of kernel-based methods [121]. Kernel regression methods frame learning and estimation problems as one of learning functions embedded in high-dimensional—or even infinite-dimensional—spaces defined over the data domain. The properties of the function space are determined by the choice of kernel, which acts as a generalized inner-product that induces a notion of distance between data samples in the function space. These types of methods have been successfully deployed in robotic systems for both dynamics and inverse dynamics learning [122–124]. However, as typically formulated, kernel methods do not have an easy way to model measurement uncertainty and noise in their function spaces. To this end, Bayesian formulations of kernel methods have been developed [125], the most common of which are Gaussian processes. Gaussian processes (GPs) are one of the primary objects of interest in the study of stochastic processes [126]. In GPs, any collection of random variables drawn from the process must be jointly Gaussian. Alternatively, one can insist that functions of the random variables be jointly Gaussian instead, which forms the basis for their application in machine learning [127]. In this context, kernels naturally arise in the specification of the mean and covariance statistics of the Gaussian process in function spaces. Using GPs, researchers have been able to parsimoniously incorporate uncertainty and noise into learning robot dynamics [128]. However, GPs, kernel methods, and nonparametric learning tools at-large typically have difficulty adapting to online learning settings such as robotic active learning. The primary underlying reason is the fact that nonparametric methods tend to grow in complexity as a function of data. Hence, as a robotic agent acquires more data it becomes more computationally expensive to make predictions with the model.

A promising compromise between the representational capacity of neural networks and the simplicity of kernel methods can be found in techniques like the Koopman operator [129]. The Koopman operator was first introduced in the study of Hamiltonian dynamics and operator theory [130]. Formally, it is an infinite-dimensional, but *linear*, operator that describes the evolution of measure-preserving dynamical systems in a lifted function space. However, to apply Koopman operators numerically they must be approximated in finite dimensions using schemes like Dynamic Mode Decomposition (DMD) [131, 132]. Algorithms like DMD use a finite basis for the function space that the Koopman operator acts on to describe the underlying dynamics [133]. Koopman operator theory and its resulting algorithms have been to a large degree developed in the context of dynamics and control, making it an ideal candidate for active learning of dynamics in robotics [134–137]. The linearity of the operator lends itself to the use of canonical control techniques such as linear-quadratic regulators, allowing for computationally-efficient nonlinear optimal control [138]. An important feature of this approach is that it does not scale in complexity with data and allows for adaptable incremental learning. The primary caveat with employing these methods is the difficulty of choosing good basis functions with which to describe the dynamics.

As an illustrative example of learning dynamics in a context that demands rapid adaptation, we compare passive and active learning in the stabilization of a malfunctioning quadrotor vehicle [115]. In this simulation, we equip two quadrotors with a data-driven model of their nominal dynamics that they can use for model-predictive control. However, at the start of the simulation we disable one of the rotors on each robot causing them to free-fall. To recover, each robot must update their internal dynamics model and stabilize themselves using control. Both agents have a single second during which they can collect data to adapt their dynamics models, after which they switch to a stabilizing controller that tries to regain control of the free-fall. Crucially, one agent learns passively and another actively by optimizing the Fisher information with respect to the unknown Koopman operator, which we discuss in the next section. Figure 3(a) shows snapshots of the different agent trajectories, indicating that the active learning agent is able to stabilize itself much more rapidly than its passive counterpart (see Figure 3(b) as well), potentially avoiding a crash. The active trajectory greatly exceeds the information gain of the passive approach (Figure 3(c)) while also achieving lower stabilization error (Figure 3(d)). Hence, by using control and movement to optimize information measures, robots can learn dynamics faster and more reliably.

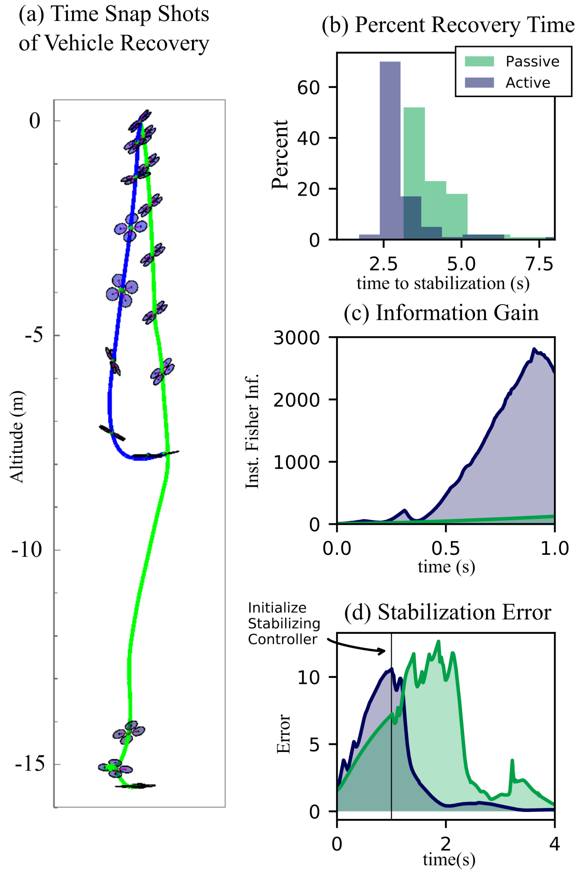

Figure 3: **Online quadrotor recovery:** Rotor vehicle recovery using active learning in a real-time single-shot learning context [115]. The rotor vehicle with the blue trajectory uses an actively learned Koopman operator representation of its dynamics. The green trajectory is the result of a passively learned Koopman operator representation of the dynamics. The rotor vehicle with the passively learned representation drops further in altitude—potentially crashing—before recovering after one rotor is disabled.

3.3. Features

The final and most broad category of learning goals we discuss is that of feature learning. Consider being blindfolded and handed a baseball and a tennis ball in either hand at random. Most people would likely be able to tell them apart with ease. But what is it about either ball that differentiates one from the other? What kinds of properties best represent each ball and its characteristics? Despite having tens of thousands of nerve endings embedded in the palm of our hand, we only need to track a few properties to be able to distinguish between the balls, such as texture and weight. We refer to the general problem of finding informative representations of high-dimensional data that can aid in a task as feature learning.

In the pattern recognition and machine learning literature, features are any measurable characteristics of a phenomenon being observed [139]. Traditional feature learning is the use of machine learning techniques to represent the "intrinsic" structure of data from raw and possibly highly-redundant measurements [140]. Recent work in this domain has focused on the use of deep learning towards finding succinct representations of human movement [141] and speech [142]. In robotics, tasks are not always well-specified and disentangling the relationship between a robot's internal state and the intended goal may be difficult. This is primarily a challenge in deep reinforcement learning where problems can become intractable when a naive state representation is used. To this end, feature learning can be leveraged towards making deep RL methods computationally tractable, and to develop schemes that better generalize to the variety of sensory inputs to which an RL agent may be exposed [143].

A simple example of feature learning can be seen in the Koopman literature. As we previously mentioned, finding the correct choice of basis functions for arbitrary dynamical systems can be very difficult. Nonetheless, recent work has been able to construct basis functions that best describe dynamics—also known as the Koopman operator eigenfunctions, or the intrinsic coordinates of the system—using deep learning [114, 135, 144]. In general, feature learning of this sort will be particularly important for robots with high-dimensional sensing modalities such as e-skins [145], or computer vision [146], and active learning can aid in enhancing rapid identification of intrinsic coordinates.

Our discussion in this review focuses on measures in Section 4 and synthesis tools in Section 5 for active learning using location and other low dimensional learning goals as examples. But the learning goal can be very high dimensional, as in the case learning dynamics of a vehicle, or in the case of learning representations (*e.g.*, machine vision applications). Regardless of whether a learning goal is low dimensional or high dimensional, the robot still has the same control authority to affect learning—it can move its body and take other physical actions to evoke response and facilitate model updates.

4. Measures for Learning

Active learning is rooted in the extraction of information from sensors [1, 94, 147–152]. Accordingly, measures of information should be expected to play a significant role. The aspects of the objective that can be captured by different information measures as well as how this information can be quantified is key in both control analysis and optimal control synthesis. The approach we discuss here follows this perspective, looking for measures appropriate both for information needs and suitable for numerical synthesis. In this section, we cover three important measures relevant to active learning—entropy, Fisher information, and ergodicity.

4.1. Entropy

Entropy-based measures have been employed in a wide range of action sensing results to calculate the expected information gain for each potential action before collecting measurements [86, 93, 101, 149, 153–166, 166, 167, 167–170]. This modern concept of entropy was developed by Claude Shannon for use in the communication and transmission of information [171]. Shannon was concerned with the amount of information necessary to reproduce the content of an information source. To this end, entropy is the expected amount of information or "uncertainty" contained in a random variable. In the case of a discrete random variable X where each x_i is a different outcome of the variable, the amount of information content in a particular event is defined by $I(x_i) = -\log p(x_i)$, referred to as bits when in base 2. The entropy of X is the expected value of the information content of each of the possible events.

$$H(X) = -\sum_i p(x_i) \log p(x_i) \tag{1}$$

The information content of a particular event decreases as the probability of that outcome increases, so low probability events provide more information than high probability events. As entropy is the average value of the information content of a random variable, the maximum value of $H(X)$ for X, would occur when each outcome of the random variable is equiprobable, *i.e.*, when there is maximum uncertainty about a particular outcome. Thus, any particular outcome for a uniformly distributed random variable does not provide much information. In the context of robotics, this is an explanation for why *rare* or *sparse* events are particularly valuable to a robot's estimation process.

By calculating the Expected Entropy Reduction (EER) of each candidate action, measures of entropy can be readily applied in the context of active sensing. However, exhaustively searching for an optimally informative solution over sensor state space and belief state is a computationally prohibitive process, as it is necessary to calculate an expectation over both the belief and the set of candidate control actions [85, 86, 101, 158, 161, 172]. Alternatively, the expected information gain can be locally optimized by selecting a control action based on a local estimate of the expected information [88, 90, 92, 94, 95, 156, 162, 166, 173]. Often times, these methods do not or cannot incorporate general sensor dynamics [88, 90, 156, 166, 173] and even the global strategies are likely to suffer when uncertainty is high and information diffuse [91, 100, 174].

4.2. Fisher Information

Active learning relies on collecting informative sensor measurements to support the learning process. In order to do so, there must be a way to locally measure the information contained in sensor readings. Used commonly in maximum likelihood estimation, Fisher information is

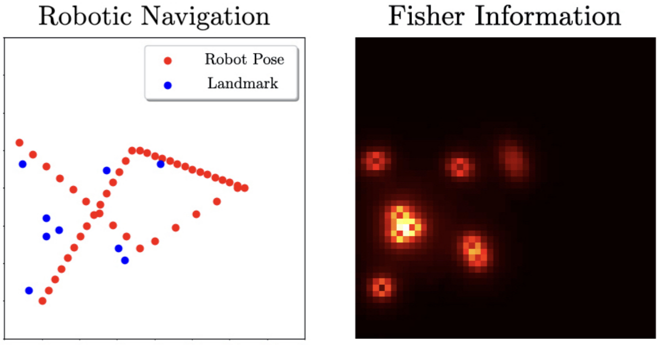

Figure 4: **Fisher information:** A measurement model indicates how a sensor will respond to the unknown parameter α, based on the current state. In SLAM applications, the measurement model might be a model of the detection of landmarks in an environment. The Fisher information distribution over landmarks provides a mechanism for determining what states the dynamical system should achieve to provide maximally informative measurements.

a method of quantifying the amount of information that a random variable X contains about an estimate of an unknown parameter, or vector of parameters $\alpha \in \mathbb{R}^M$. Using $p(x|\alpha)$, the density function of X parametrized by the value of the vector α, one can determine the likelihood of an observation x given a value of α. The Fisher information is an $M \times M$ matrix that captures the local sensitivity between parameters and observations [170, 175]:

$$\mathcal{I}(\alpha) = E_X\Big[\Big(\frac{\partial}{\partial \alpha}\log p(x|\alpha)\Big)\Big(\frac{\partial}{\partial \alpha}\log p(x|\alpha)\Big)^\top\Big|\alpha\Big], \quad (2)$$

where the expectation is taken over realizations of X at a given value of the parameter vector α. If $p(x|\alpha)$ is highly sensitive to changes in α—*e.g.*, the distribution of observations exhibits a steep dependence on α—then for a given measurement there will be a small range of highly probable values of α. If $p(x|\alpha)$ is not sensitive to changes in α, then there will be many candidates of comparable likelihood.

In robotics, Fisher information is well suited for measurement models that are naturally parametric (*e.g.*, size, weight, location). Measurement models, sometimes called observation models, are predictions of how unknown variables will impact a sensor reading. This sensor reading can be very sophisticated, like a camera being used in a pixels-to-torque application [176], or very simple, such as a one-bit sensor being used for trajectory tracking [177, 178]. The measurement model provides a way of expressing *what* the robot is attempting to learn in terms of its sensing capability and means to adjust its sensors. A commonly used measurement model form is $z = \Upsilon(\alpha, x) + \Delta$, where z is the measurement, α is the parameter being estimated, x is the state of the agent, and Δ is (possibly multi-dimensional) zero-mean Gaussian noise. This model is in the form of a sum of a deterministic term—typically modeled by first-principle physics—and a noise term which can be rather challenging to justify, since most robotic applications will not have such convenient additive normal distributions.

For active learning applications, measurement models can play an important role in calculating information measures over a space. To estimate a parameter vector α, the Fisher information matrix has each element (i, j) given by:

$$\mathcal{I}_{i,j}(x, \alpha) = \frac{\partial \Upsilon(\alpha, x)}{\partial \alpha_i}^\top \Sigma^{-1} \frac{\partial \Upsilon(\alpha, x)}{\partial \alpha_j}, \quad (3)$$

where the multi-dimensional noise is assumed to be zero-mean Gaussian with covariance Σ. Intuitively, Fisher information can be expected to be higher where the expected measurement signal is greater than that of the noise. The expected information density $EID(x)$ over a search space can be constructed by computing the *expected* Fisher information with respect to a probability distribution representing an estimate of a parameter $p(\alpha)$. This $EID(x)$ would then form the information landscape against which active learning decisions are made and then executed.

As an example, we consider the use of the Fisher information in Simultaneous Localization and Mapping (SLAM) problems subject to measurement models of the form discussed above. While the SLAM literature in robotics is diverse and well-established, the more recent field of *active* SLAM has seen much growth [179]. Active SLAM makes use of representations of uncertainty and information to generate exploration plans. In active SLAM, different information measures can capture different features of an environment. In Figure 4, measurement models for landmark detection are used to provide a basis for calculating information measures to inform the agent's exploration plan. In this case, the Fisher information over each landmark attracts the robot to landmarks with lower uncertainty, thereby enabling efficient loop closure. This allows an agent to discern an ensemble of locations that are expected to provide more informative measurements.

4.3. Ergodicity

Ergodicity is a fundamental property of dynamical systems and stochastic processes. Formally, achieving ergodicity implies that the dynamical system uniformly visits all parts of the space in which it exists [180]. However, more often what we mean when we say that a system is "ergodic" is whether or not it satisfies the pointwise ergodic theorem [181]. In this sense, being ergodic requires that the system spend time in regions of space in proportion to the measure of said regions. The specific measure used can vary with context, but very often probability measures are used.

In engineering contexts such as active learning, we are free to choose or construct the spatial measure. Particularly, when a system is ergodic with respect to measures representing an information distribution over the space, ergodicity demands perfect asymptotic sampling of informative states. As a simple example, consider a system trajectory $x(t) \in \mathcal{X}$ and a probability density function (PDF) capturing the expected distribution of information over the space. If the trajectory is ergodic, then the amount of time

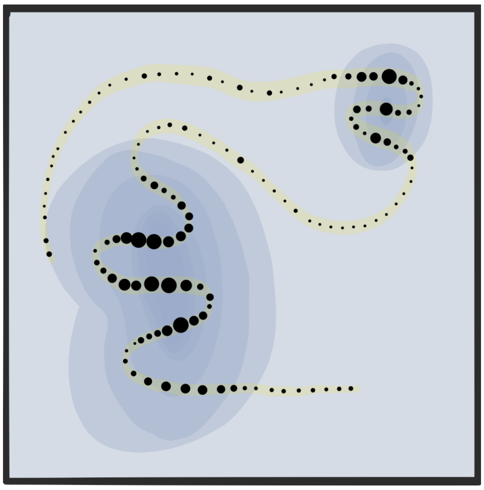

Figure 5: **Ergodicity:** For an agent to be ergodic with respect to a target distribution, the spatial statistics of the agent's trajectory must match the statistics of the target distribution. This means that the time spent in a particular area is proportional to the density of the target distribution in that area. Here an agent traverses a bimodal distribution. The size of each waypoint is proportional to the time spent at that location.

the agent spends in each neighborhood $\mathcal{N} \subset \mathcal{X}$ is going to be proportional to the amount of information in $\mathcal{N}$ as measured by the PDF (see Figure 5). Hence, designing ergodic dynamics with respect to desired measures is of interest to active learning [182]. However, in order to do so we need a metric that captures how "ergodic" our trajectories are.

Because perfect ergodicity is only possible on infinite time horizons, we require a metric that can be maximized over finite-horizons through decision-making—such a metric was developed in [183]. Metrics on ergodicity provide a principle of motion [13, 24] similar to energy minimization and error minimization, and can be used to synthesize automated exploration for learning, as we will see in Section 5. The ergodic metric in [183] provides a method for comparing a trajectory $x(t)$—a singleton at any given time t—to a distribution $\Phi(x)$ through their *spatial Fourier transforms*. This suggests that one can compare the coefficients c_k of $x(t)$ and ϕ_k of $\Phi(x)$ respectively and measure a distance between the two. In general, it is not obvious how one might do this otherwise since information content between dimensionally different objects is typically not well defined.

Comparing how "close" two quantities are to each other is imperative for control when using optimization-based methods. To compute the Fourier coefficients ϕ_k of a distribution $\Phi(x)$, we use the inner product

$$\phi_k = \int_X \phi(x) F_k(x) dx, \tag{4}$$

where F_k's represent the choice of Fourier basis functions. For trajectories, we begin by interpreting them as distributions comprised of sequences of impulses:

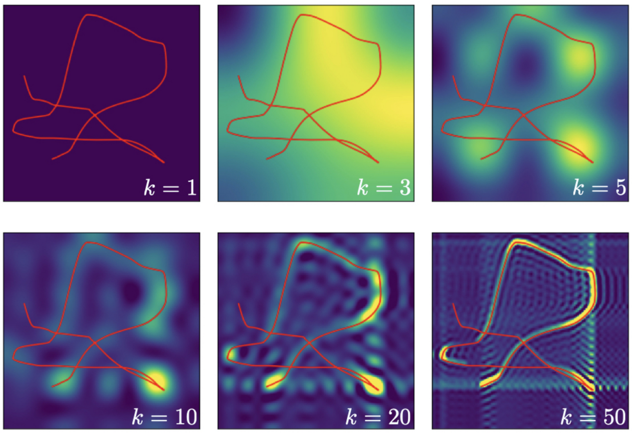

Figure 6: **Fourier transform of a trajectory:** The Fourier transform of a constant speed trajectory represents the trajectory in the form of a spatial distribution. The representation of the trajectory by its transform changes in granularity for $k = 1, 3, 5, 10, 20, 50$ Fourier coefficients.

$$C(x) = \frac{1}{T} \int_0^T \delta\left[x - x(t))\right] dt, \tag{5}$$

where δ is the Dirac delta [183]. Then from the properties of the Dirac delta function, we can calculate the Fourier coefficients

$$c_k = \frac{1}{T} \int_0^T F_k(x(t)) dt, \tag{6}$$

where the coefficients take on the value of the basis functions averaged over a time window of duration T. An example of such a spatial representation is shown in Figure 6, where a trajectory along with its Fourier decomposition is shown for different numbers of coefficients c_k. As the number of coefficients k increases the spatial resolution of the trajectory improves, showing how the statistics of a trajectory may be represented as a spatial distribution.

With this in mind, the ergodic metric represents a *distance from ergodicity* that is measured from a time-averaged trajectory $x(t)$ with respect to a distribution $\Phi(x)$. This distance is calculated by imposing a norm on the difference between the trajectory's c_k and the distribution's ϕ_k coefficients. Particularly, we take the Sobolev norm between the coefficients by using the sum of the weighted squared distance between them:

$$\mathcal{E}(x(t)) = \sum_{k_1=0}^{K} \cdots \sum_{k_n=0}^{K} \Lambda_k \left|c_k - \phi_k\right|^2 \tag{7}$$

where K is number of Fourier coefficients used for each of the n dimensions, and k is a multi-index $(k_1, ..., k_n)$. The coefficient $\Lambda_k = (1 + ||k||^2)^{-s}$ is a weight where $s = \frac{n+1}{2}$, which places larger weight on lower frequency information, ensuring convergence [183]. It is worth noting that spectral methods, and the ability to generate a norm on a trajectory $x(t)$ using them, offer opportunities in measuring entropy as well. The entropy of a distribution could also be measured in the Fourier domain—yielding an objective function that is differentiable and amenable to control synthesis, enabling one to avoid approximating entropy in an optimization with Fisher information (e.g., as in [184]).

The measures discussed in this section form the basis for how we measure performance of an active learning system. The next section focuses on synthesizing decisions that optimize, or at least improve, those measures.

5. Control Synthesis for Active Learning

Active learning has a wide range of applications in robotics including prioritized decision making [159, 185], inspection [165], mine detection [186], object recognition or classification [155, 160, 187], next-best-view problems [156, 157, 188], and environmental modeling [97, 98, 189]. As a result, particular controller architectures may be advantageous for different environments, tasks, and constraints. Here, we survey several model-based optimal control methods that provide distinct advantages for active learning.

Model-predictive control (MPC) is an optimal control framework that optimizes current actions with respect to an objective while taking into account the future behavior of the system over a finite time horizon. Once the current action is taken, MPCs reoptimize from the new starting point and continually plan actions throughout the receding horizon. MPCs are particularly suited to active learning because receding horizon planning lends itself to continuous incremental learning, while simultaneously enabling assessments of the safety and stability of trajectories. In contrast, other optimal control approaches such as the linear-quadratic regulator (LQR) must solve the entire control problem without replanning.

One of the primary optimal control algorithms is Differential Dynamic Programming (DDP) [190], which is an extension of the seminal work by Bellman [191]. DDP is a model-predictive method requiring second derivatives of the dynamics to realize quadratic convergence to the optimal solution. While DDP has fast convergence guarantees, calculating the Hessian of the dynamics can be computationally intractable. If one is willing to forego the fast convergence rate by disregarding the second order terms of the control solution, DDP becomes equivalent to the first order iterative LQR (iLQR) method. DDP and iLQR have both been shown to be effective in the context of robot control in a variety of applications [192]. For example, in [193] the authors use local trajectory optimization methods in combination with RL to learn policies for dexterous manipulation with a five-fingered robotic hand. In scenarios where the dynamics are known or easily modelled, and their Jacobians and Hessians are inexpensive to compute, DDP and iLQR may be well-suited to active learning applications.

A method that generalizes MPC to both convex and nonconvex objectives is the sampling-based Model Predictive Path Integral (MPPI) control algorithm [194]. In MPPI, Monte Carlo sampled trajectories are used to approximately extremize a free energy objective [195]. These types of objectives are designed in analogy to thermodynamic free energy from the statistical mechanics literature and can be used to synthesize control [196]. Moreover, the synthesized control actions are formally equivalent to Bellman optimal control without the need for computing derivatives, and their computation can be easily parallelized [197]. As a result, MPPI is particularly well-suited for use in learning problems where the dynamics of the agent are non-differentiable or too complex to differentiate in a computationally-efficient way as with neural network models. For example, in [194] the authors use MPPI to learn a neural network model of the dynamics of an auto-rally autonomous race car. However, depending on the structure of the task, generating enough simulated trajectories to sufficiently sample a learning objective may become prohibitive.

Another model-based control synthesis method is Sequential Action Control (SAC), which is inspired by hybrid systems theory [198]. Unlike other MPC techniques, SAC explicitly tries to expend the least control effort possible in generating actions by taking into account the benefits of taking optimal actions as opposed to alternative policies or doing nothing. SAC simultaneously finds the actions that optimize an objective, the best time to apply said actions, and the application duration. Due to its hybrid specification, SAC can naturally handle non-smooth dynamics, and can also be easily wrapped around other controllers to enable more exotic control architectures [199]. In [200], SAC was used for active parameter estimation with a robotic system. This work uses SAC to control a robot to determine the length of a pendulum by maximizing the Fisher information with respect to the pendulum parameters. The SAC control actions sequence is piecewise continuous, with generally short application durations for any control. This allows a robot to reactively generate motions towards information dense regions. However, like most MPC techniques, it requires having access to the derivatives of the objective and dynamics, which can constrain its usage in learning scenarios as previously discussed.

An important consideration when choosing a controller for active learning is the global characteristics of the search process. Depending on the structure of the learning task, there may be a single optimum that represents the true parameter value that is being estimated. Other learning tasks require that the agent avoid fixating on a single information source and instead visit many sources. We distinguish between these approaches by referring to them as myopic and non-myopic respectively. Myopic learning uses local algorithms that greedily take actions over short horizons that optimize the immediate learning objective. While these methods are prone to getting trapped in local minima, they have lower computational overhead than non-myopic learning methods. Non-myopic approaches plan control actions over long time horizons so as to pro-

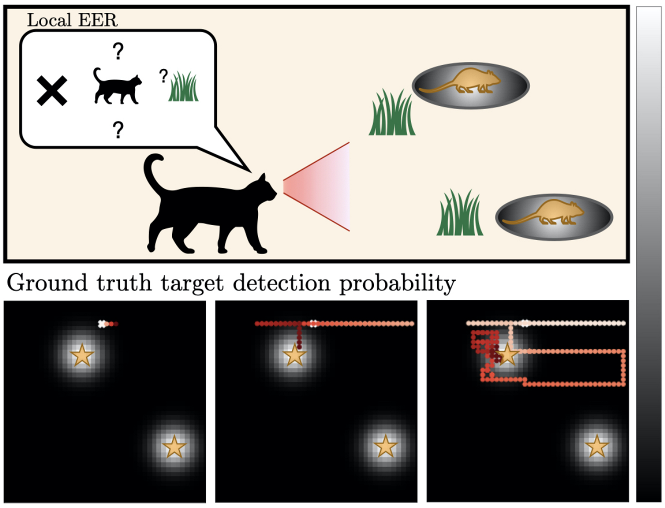

Figure 7: **Infotaxis:** *Upper panel:* A cat searches for the two yellow mice. The distribution around the mice represents the probability of detecting the mouse at that location where white is high probability. The gradient in the cone around the cat represents the cat's measurement model. The cat has a mental image of the expected local reduction in entropy for moving up, down, left or right. *Lower panel:* This is an example of an infotactic trajectory of an agent searching for target locations, represented by the yellow stars. The likelihood of detecting the target is represented by the distribution around the target locations, becoming more dense closer to the target. The measurement model encodes the ability of a camera to detect an object at a particular range and angle of attack. The search strategy selects the direction of movement that maximizes the expected entropy reduction at each time step. The infotaxis strategy succeeds at finding only one of the two targets in the environment and stops searching.

duce coverage over distinct information sources. These are often used to avoid local minima associated with fixation [98, 100, 101], and can take advantage of approximate solutions [86, 87, 91, 98–100, 164, 186, 201, 202].

Choosing a mechanism by which one can avoid myopic learning is critical to operating in environments that have unmodeled effects, such as visual occlusion, where the *expected* most informative state may not provide information. For example, a camera taking a picture of a person behind an a piece of furniture does not benefit from multiple pictures taken from the same state. As a result, dynamic coverage of high information density areas can keep a robot collecting good data while acknowledging unmodeled uncertainty effects through decision-making. Taking these factors into account can be critical to the success of the active learning process. Next, we will examine two approaches to active learning and exploration—infotaxis and ergodic control—that take opposing attitudes towards this question.

5.1. Infotaxis

Inspired by animals' search for chemical sources in a fluid such as air or water, infotaxis is an information-maximizing search strategy using entropy reduction as an information criteria [153]. This technique was developed to show that a search plan does not need to depend on environmental gradients, such as the concentration of a scent smoothly increasing in proximity to a flower. Instead, animals may sense traces of a source dispersed by wind or currents and formulate a movement strategy based on infrequent detections.

In this work, an agent attempts to localize a target or source in a 2D environment based on detections of the target. To generate an infotactic trajectory, the searching agent chooses a control action at each time step that locally maximizes the expected reduction in entropy, thereby maximizing expected information gain. Concretely, the agent considers moving to adjacent positions on a lattice, or staying in the same location to take more measurements.

To determine an action, it is necessary to have a probability distribution $p(r)$ representing the unknown location r of the source. The probability of detecting the source at a given location is dependent on the distance from the source, meaning that the record of detections along the trajectory of the searcher, $x(t)$, carries information about the source location. When a detection event occurs, the times and coordinates are stored in the random variable $\mathcal{T}_t$. From this record of detections, the searcher is able to represent the location of the source as a posterior probability distribution that is updated based on the measurement taken at each time step.

$$p_t(r_0) = \frac{\mathcal{L}_{r_0}(\mathcal{T}_t)}{\int \mathcal{L}_x(\mathcal{T}_t)dx} \tag{8}$$

Here, $\mathcal{L}_{r_0}$ is the likelihood of detections $\mathcal{T}_t$ for a source located at r_0. From the posterior distribution one can calculate a control action that minimizes the expected entropy at the next time step by selecting a set of potential actions, computing the EER given the current $p(r)$, and then selecting the action that provides the minimal EER. This strategy can be computationally prohibitive for many systems.

The trajectories produced by infotaxis exhibit similarities to biological organisms such as moths or bacteria that engage in olfactory search [203]. However, infotaxis-type approaches can fail when there are distractors—states that appear similar to the target but are not the target—in the environment [13]. The searcher may conflate the actual target with the distractor and then ignore the intended target. Practically, infotaxis can only be implemented using short time horizons as the computational requirements of predicting for longer horizons are significant. For each control action considered, the expected entropy reduction must be calculated, including calculating a posterior for each possible outcome of the measurement random variable. Figure 7 provides an example of an infotactic search with two target locations. Here, the agent successfully determines the location of one source and stops searching. This strategy is purposefully ignorant of a signature that may conflict with the perceived location of the target in favor of detecting the same target to increase its cer-

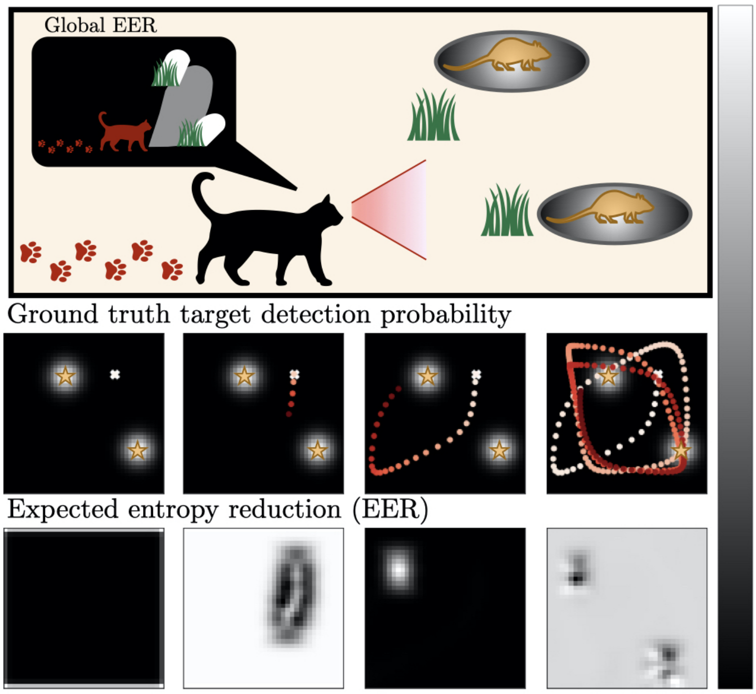
Global EER
Ground truth target detection probability
Expected entropy reduction (EER)

Figure 8: **Ergodic control with respect to the expected entropy reduction over the search space:** *Upper panel:* A cat searches for two targets represented by the yellow mice. Here, the cat has access to its past trajectory and has a mental image of the expected entropy reduction over the whole search space. *Middle panel:* An agent is searching for two targets located at the yellow stars. The information distribution becomes more dense closer to the targets. Here the agent takes a measurement and updates its belief using the same measurement model and likelihood function as in the infotaxis implementation. *Lower panel:* The agent chooses its next control action based on the global expected entropy reduction. This is determined from its belief of the information content in a particular location.

tainty. This example illustrates that the infotactic strategy is myopic when confronted with multiple sources or environments with convincing distractors.

While an infotactic search strategy can experience difficulties when there are multiple targets in the environment that require persistent monitoring, it is well suited to react to sporadic cues and requires only local information. Infotaxis represents one of the most straightforward examples of active learning in which an agent acts greedily to maximize expected entropy reduction.

5.2. Ergodic Control

Recent work by the authors and colleagues has analyzed biological motion by introducing energy constrained proportional betting [13, 204], where the energetic cost of movement is balanced against the desire to gain sensory information about a source. This approach uses the ergodic metric, discussed in Section 4.3, to quantify how well a trajectory covers a distribution of expected entropy reduction. The resulting algorithm produces trajectories that balance informative sampling—collecting many samples in high information areas—with the amount of energy expended from motion. These types of trajectories were observed in the behavior of electric fish, moles, and cockroaches. This suggests that the strategy of energy constrained proportional betting provides a competitive hypothesis for the ways in which living creatures collect information about their surroundings, and may be a robust approach for robotic systems to acquire information. Extensions and variations of this idea now arise in many robotic applications [205–213].

If the goal of infotaxis is to maximize the information content of a series of measurements collected along a trajectory, the goal of ergodic control—first developed in [183]—is to control the spatial statistics of a trajectory $x(t)$ to match those of an expected information density distribution $EID(x)$. This requires the choice of a norm on the difference between the distributions $EID(x)$ and the trajectory $x(t)$ interpreted as a distribution $C(x)$, defined in Equation (5). To this end, we use the ergodic metric from Section 4.3 as an objective to synthesize maximally ergodic trajectories for general nonlinear systems using tools from model-predictive control [204]. However, we note that any trajectory optimization tools or direct optimization tools could be used; we use the results from [204] primarily because they are amenable to real-time computation [214].

The first thing to note is that the ergodic metric $\mathcal{E}$ in Equation (7) is not of the form of a running cost—as a result it is not a Bolza problem (although one can turn it into a Bolza problem by appending the Fourier coefficients to the state vector [215], creating an infinite dimensional state space). Nevertheless, one can calculate the adjoint variable ρ of the ergodic metric function:

$$\dot{\rho} = -\frac{2}{T}\sum_k \Lambda_k \left(c_k - \phi_k\right)\frac{\partial F_k}{\partial x} - \frac{\partial f}{\partial x}^\top \rho \tag{9}$$

where the dynamics are represented by $\dot{x} = f(x, u)$, and get a descent direction for locally minimizing the ergodic metric [216]. Other approaches can be used that lead to slightly different solutions (*e.g.*, the projection-based trajectory optimization method for ergodic control in [204, 217, 218], where higher-order convergence properties come at the expense of high computational cost). A key property of the metric $\mathcal{E}$ is that it is differentiable with respect to $x(t)$, so most optimal control techniques can be easily applied.

An example of an ergodic trajectory can be seen in Figure 8, where the agent is exploring with respect to the expected entropy reduction distribution over the whole environment. The agent is able to successfully locate both target locations in this scenario because the ergodic control strategy is amenable to persistent monitoring of multiple targets. As perfect ergodicity can only be realized as time goes to infinity, the agent will continue to explore the space. Using infotaxis, the agent would conclude its exploration once a target has been detected. Here, we make use of global information to plan control actions over longer time horizons.

With both these local and global information-based synthesis techniques in mind, we next move on to applications in robotics that will depend on active learning strategies.

6. Applications in Robotics

While the landscape of applications for active learning is almost as broad as that of machine learning itself, here we will focus on settings where datasets are rarely available ahead of time. Active exploration applications such as search and rescue or mapping are particularly relevant in this class of problems, especially when the environments are dynamic and hard to predict. We also discuss applications in which system models are either unknown or difficult to parametrize, as is the case for soft robotics and for many of the areas of application of imitation learning.

6.1. Soft Robotics

Soft robots are made from compliant materials, enabling them to be well suited for delicate tasks and environmental adaptation [219–221]. Unfortunately, precise modeling and control of soft robots poses challenges because soft materials are continuously deformable and thus nominally have infinite degrees of freedom. There is no clear method of representing the geometry of such a robot without making significant simplifications [20]. The most important functional property of a soft robotic system—deformation in response to the environment—makes soft robotic systems practically impossible to meaningfully model for control based on first-principles (*e.g.*, partial differential equations based on elastic body mechanics). Data-driven modeling is a natural alternative when first principle arguments are either not tractable or do not involve the use of a state space.

Learned representations, such as those constructed by DNNs, have been shown to find input to output mappings that predict the behavior of soft robots [222]. However, these models are difficult to apply using known model-based control techniques. Alternatively, the Koopman operator has also been used for modeling and control of soft robots [137]. Described earlier in Section 3, Koopman operators provide a linear representation for nonlinear dynamical systems that is compatible with linear control methods such as LQR synthesis. In practice, a data-driven approximation is adopted. As an example, [137] develops a model predictive controller with a Koopman operator representation of a soft robotic arm for tracing reference trajectories. The data collection strategy for soft systems plays an important role in determining a model. For instance, though obvious, data collected while an end-effector is out of contact with the environment cannot provide useful modeling data. In prior work we showed that a Koopman operator representation of a robotic system can be actively learned using information-theoretic strategies [114].

Despite the complexity that soft elastic structures introduce to the analysis of robotic motion, soft robots can beneficially exploit these physical properties. For example, soft structures can be leveraged as a computational resource, sometimes called morphological computation or embodied intelligence [223]. A soft body that deforms around an object, in principle, will make manipulation easier, and will imply that the amount of explicit computation needed will be lower in exchange for the implicit computation afforded by the soft body. For instance, [224] shows that stable hopping behavior of a soft underwater robot can be achieved experimentally by dynamically changing the size of its body. Moreover, with actuator saturation, adapting the morphology of the robot's body was the only route to achieve stable behavior, implying that control over the continuous shape properties of the robot was key to task success.

In addition to articulation, sensory acquisition via morphological computation is connected to biological systems and present in structures such as the cochlea of the human ear [225] and the bodies of octopuses [226]. While data can be passively collected through the physical structure, active sensing is a biologically motivated extension. In [120], the authors build a perception system to learn the kinematics of a soft actuator and estimate interaction forces with embedded sensors and recurrent neural networks. In their approach, the authors consider the relationship between action and perception in the learning process by quantifying sensor information as a result of commanded actuation information. Work in [227] uses a soft robotic probe to palpate imitation tissue to determine the location of a hard tumor-like nodule. The soft robot was able to adjust its stiffness across iterations of the palpation task based on information metrics calculated from human test subjects. These findings suggest that active haptic perception through physical changes to the probe improves estimation accuracy, motivating active learning techniques that could automate learning for this and other soft systems.

6.2. Search and Rescue

Prevention, response, and recovery from disasters can be dangerous for emergency professionals who may need to interact with areas affected by events such as hurricanes, oil spills, and earthquakes. Disaster robotics is an area that works to augment the capabilities of workers by delivering real-time data to experts and intervening in the environment [228]. The need to efficiently search an environment is an issue at the core of disaster robotics. One of the most visible examples of the need to search an extremely large, dynamic environment in recent years is the investigation of the crash site of Malaysia Airlines Flight 370 (MH370) in March of 2014. In the first 52 days after the crash, the Australian government reported that air crafts and surface vessels covered an area of over 1.6 million square miles. By June of 2018, the final search effort was suspended without success. Although there may be many points of failure in this search effort, one dimension involved robotic technologies that scanned the bottom of the ocean that were incapable of reasoning about potential debris signatures, the dynamic environment, and their own capabilities.

When searching large areas where information is sparse, active coverage algorithms are important in determining important areas of a search region and the schedule to visit these regions. Coverage algorithms are used in many robotic applications such as underwater exploration [229], agriculture [230], and inspection [231]. The goal of coverage algorithms is to visit all points in an area or volume while avoiding obstacles [232]. Commonly used approaches for coverage, a taxonomy of which is included in [233], include cellular decomposition or grid-based methods to divide the area into manageable sections [234–237]. However, as the complexity of the environment increases, the number of cells necessary to represent the environment increases. These methods typically do not take into account the physical properties of sensing capabilities of the robots or the dynamics of the environment. As a result, coverage is treated as both necessary and sufficient for capturing needed data. This attitude about coverage can be seen in the search strategy of the MH370 investigation which focused on area coverage, neglecting factors such as how the ocean currents might pull debris away from the site [238].

6.3. Localization and Mapping

SLAM algorithms create a map of an unknown environment while simultaneously estimating the state of the robot within that environment. This is a major success story in robotics, with the current flood of driverless car technologies all dependent upon SLAM algorithms. When navigating an unknown environment, a robot may lose its ability to localize itself due to accumulated small errors in sensors and actuators, known as representation drift. To correct for this drift, SLAM algorithms use loop closure—the task of identifying whether an agent has returned to a previously visited location—to maintain an accurate representation of the location of the robot relative to environmental features. To maintain loop closure, the robot revisits regions with low estimation uncertainty or informative features to combat representation drift. Beyond localization, loop closure allows the robot to represent the topology of the environment, instead of simply a record of where it has been.

In passive approaches, a robot performs SLAM with sensor information provided to it. For instance a lidar sensor collects data while driving down a road. In contrast, active SLAM leverages the actions of the robot to seek out informative measurements that efficiently decrease localization and mapping uncertainty. Figure 9 illustrates the flow of information in passive versus active SLAM. Active SLAM generates controls based on the current state of both the map estimate and robot states. The review paper [179] summarizes methods that have been employed in the development of active SLAM including the theory of optimal experimental design [239], information theoretic approaches [240–242] and control theoretic approaches [243, 244]. Active SLAM can also be formulated as a Partially Observable MDP (POMDP) and approximated using Bayesian optimization or Gaussian belief propagation to attain computational tractability. Belief space planning entails planning in the space of probabilistic estimates of a robot's state and additional variables of interest [245, 246]. This method has also been used in combination with navigation error [247–249].

Using planning algorithms in SLAM is challenging because SLAM is generally executed on a pre-planned trajectory. This trajectory can greatly affect the quality of performance. Conversely, path planning algorithms typically assume a given map. Hence, planning and SLAM are nontrivially interdependent. Work in [250] attempts to integrate SLAM with a coverage path planning problem by developing a movement strategy they call perception-driven navigation. The authors use a cost function that weights navigation uncertainty, evaluated using the Fisher information matrix described in Section 4, with the ratio of unexplored to total coverage area. This method plans paths between waypoints that are selected based on a measure of visual saliency, prioritizing areas in which notable environmental features have been detected. The integration of perception based navigation in the SLAM framework is key to balancing effective mapping alongside exploration as the distribution of features in an environment is often highly uneven. It also allows for operating in limited field of view environments, such as underwater

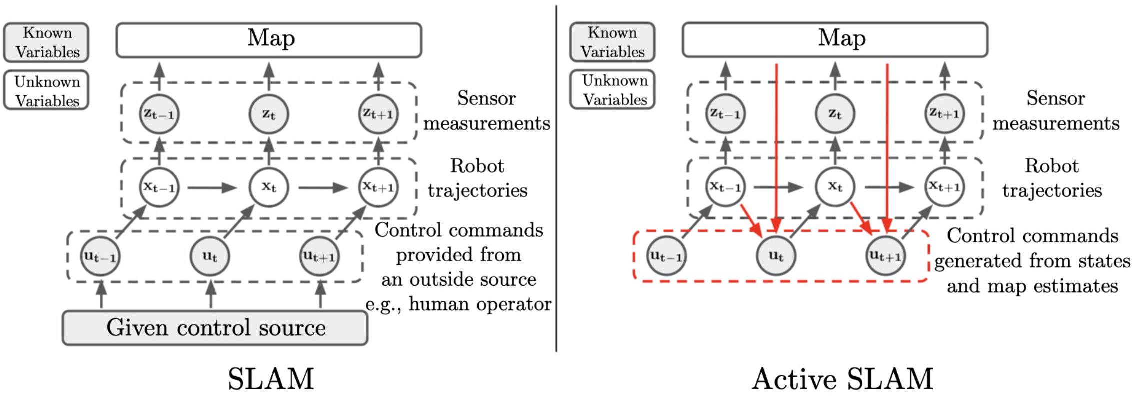

Figure 9: **SLAM versus Active SLAM:** Active SLAM uses control commands generated to decrease localization and mapping uncertainty. In traditional SLAM, the control signal is given in the problem statement.

inspection tasks.

Developing a method to determine informative features from images is an important aspect of visual SLAM, in which SLAM is performed using only camera inputs [251]. Image pre-processing with feature selection reduces the computational burden of scanning all the pixels in images, leading to many active feature selection algorithms [252]. One can also selectively process informative regions of images or videos using a recurrent neural network [253]. Lastly, visual-inertial navigation—where a robot must estimate its state using only a camera and inertial sensors—can supplement the visual SLAM process. In [254] visual-inertial navigation selects features based on the state of the observer and the context of the scene, using information theoretic constructions as a basis for prioritizing features to be used in state estimation.

6.4. Imitation Learning

Imitation learning is a widely used and effective method of imparting human skills to robots by learning desired behaviors from demonstrations. To transfer knowledge about a task through imitation, it is important to capture salient features of a demonstration in efficient and generalizable representations of a skill. Here, active learning can play an important role in capturing knowledge from a demonstration.

The field of imitation learning is expansive [255–257] and has been used in numerous settings including autonomous driving [258], virtual games [259], and replicating human motion in robots [260]. Capturing knowledge about a task from human experts is especially applicable to robotics, where autonomous systems are charged with operating in complex and unstructured environments. In these situations it can be difficult to manually program specific behaviors and engineer reward functions to suit a task. Imitation learning is commonly tied to deep neural networks to take state/action pairs from demonstrations and learn a policy for a skill. This can often require large amounts of data, leading to questions about what aspects of demonstrations are particularly useful to impart a skill to an autonomous system.

When transferring skills from a human operator to a robot, active learning occurs when a human operator is queried for information. For instance, work in [261] considers two approaches to active learning from demonstration in the context of autonomous navigation. A learner, such as a robot, selects expert demonstrations that they believe to be informative based on either novelty or uncertainty reduction criteria. In novelty management, demonstrations are selected based on a density model from which a test feature vector can be compared to demonstrations previously seen in training to provide exposure to unobserved or anomalous data. For uncertainty reduction based active learning, the authors used the Query Bagging Method [262], in which training data is partitioned into multiple subsets. A demonstration would be deemed to have high uncertainty if the variance over these subsets for the demonstration was high.

Inverse reinforcement learning (IRL), also called inverse optimal control, is a method of determining the goals of desired behavior from trajectories executing a policy [263]. The aim of IRL is to find a reward function that describes the desired task from expert demonstrations. When a task is well suited to be described by a single reward function, IRL is most applicable. However, a policy may be optimal for multiple reward functions, making it difficult to discern intent. In response, it may be necessary to include other objectives. Work in [264, 265] focuses on active learning in the context of IRL, which seeks to reduce the demonstrations from full trajectories to particularly useful states. In this case, active learning means selecting particularly informative samples to be labeled by an oracle. In [264], a robot learns a reward function and movement policy for a grasping task. The reward function is in the form of a Gaussian process model and is based on human evaluations of the quality of the grasp. In this method, the learning agent is able to impact the demonstrations it sees by choosing to query human expert ratings based on acquisition functions from the Bayesian optimization literature.

Generative adversarial imitation learning (GAIL) is a model-free imitation learning approach that scales well to high dimensional environments [266]. Inspired by generative adversarial networks, GAIL produces behaviors similar to demonstrated behaviors while training a discriminator to differentiate expert attempts with generated attempts. An extension of GAIL, called InfoGAIL, attempts to find latent structure across human demonstrations—that can be highly variable—to describe interpretable concepts [267]. Related to techniques that train a discriminator to differentiate between expert and learned policies(such as InfoGAN [268]), InfoGAIL approximately maximizes mutual information between latent space and trajectories to deduce meaningful latent variables. In this way, it is possible to produce semantically meaningful or informative data that pertains to a particular task.

Imitation learning, and the other applications mentioned above, stand to benefit from robots that physically manipulate when and how they learn, rather than relying on visual and aural requests for more or better data, which is one of the principal goals of active learning in robotics.

7. Open Challenges

Closed-loop active learning presents a key opportunity for improving the quality and rate of learning. In this section, we focus on specific challenges in both the near and far term, such as safety and distributability. These challenges are specific to the expertise of the controls community—*e.g.*, analyzing properties like complexity, convergence, and motion feasibility. We end with a broader discussion of questions such as how can one assess the sufficiency of a learning model for a given task? These challenges, among others that we may not yet understand, are at the core of what it means to construct a robotic theory of active learning.

7.1. Distributability

Distributability has become a widely studied and often implemented goal for control systems, enabling a swarm of robots to accomplish what an individual robot cannot. In the context of control-driven tasks such as exploration or search, the benefits of distributability are immediately apparent—multiple robots will be able to cover an area more efficiently than a single robot could. Distributed data collection of this kind has been widely and successfully applied in a variety of contexts, such as environmental monitoring [236, 269]. The key feature underlying the success of these distributed control applications is that the dynamics of the robot collective are factorable into a block-diagonal representation—the dynamics of each robot agent are independent from one another [214, 270]. However, can we expect this to be the case across active learning applications?

While independent robots can easily coordinate to collect measurements and effectively augment their perception [271], learning collectively may prove to be much more challenging for a variety of reasons. For one, when robots are not just collecting data but also using it to learn as a group, they must be in constant communication and sharing data samples with one another. Another important challenge is that the data samples that each agent is locally exposed to may be statistically distinct. Moreover, the noise and disturbances that robots are exposed to may be heterogenous across agents as well. Taken together, these observations suggest that during distributed learning the samples that a swarm collects may not be independent and identically distributed, which is a key assumption underlying most learning methods and can create issues with fundamental properties of the learning process (*e.g.*, convergence). Most of the difficulties outlined so far have been described by the fields of distributed [272] and federated [273] machine learning. Hence, the success of distributed active learning is in part tied to the challenges of distributed learning generally.

Nonetheless, some challenges in distributability will be unique to active learning. As we have discussed, when the dynamics of robotic agents are left uncoupled making control decisions may be simple. However, active learning in robotics precisely requires a coupling between learning and taking actions. Then, when agents share a common distributed learning objective, their dynamics may become effectively coupled through the contingent relationship between acting and learning. As a best-case scenario, this can lead to redundant data collection and learning, but in the worst-case this can create stability issues in the learning process. Highly-coupled dynamics, along with extended network dropouts, will generate high degrees of disagreement between agents, making both analysis and prediction more difficult. Thus, eliciting useful collective behavior from decentralized systems based on local decisions is still an open challenge.

7.2. Safe Active Learning

Safety is a problem of both specification and prediction—one needs to specify what is meant by safety and be able to predict that the specification will be satisfied. Imposing safety enables learning in high-consequence environments with continuous deployment, making reliance on models and prior experience less risky.

Common tools available for imposing safety constraints often depend on Control Lyapunov Functions (CLFs) [274, 275]. These control approaches enforce stability properties of a robotic system through a feedback stabilizing control law that drives a positive-definite differentiable function to zero over time. In the context of active learning, one may desire to have a CLF for ergodic control, using the ergodic metric as the candidate Lyapunov function [216]. One can use Control Barrier Functions (CBFs) [276–278] that encode safety constraints, such as in Figure 10 where we impose the constraint that one set of vehicles can only

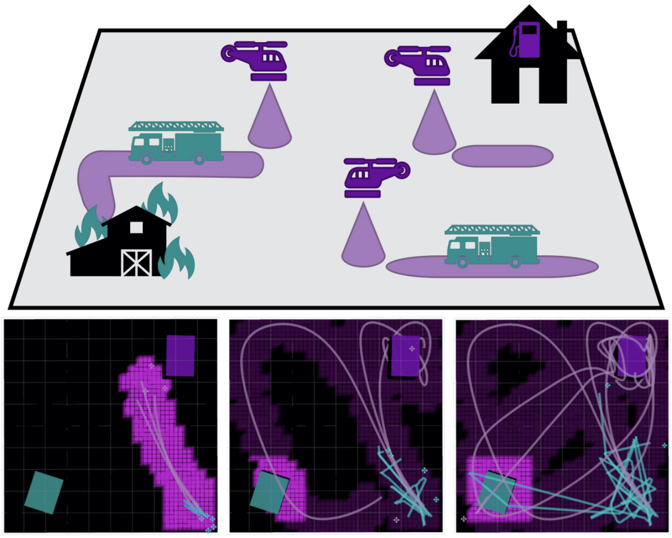

Figure 10: **Safe decentralized ergodic control:** *Upper panel:* Fire trucks attempt to reach the site of a fire guided by helicopters above. The firetrucks are able to explore in the areas that have already been explore by the helicopters. At the same time, the helicopters must maintain the ability to return to a refueling station. *Lower panel:* Here are three time snapshots of an ensemble of six robots—three purple and three blue—explore an environment subject to the condition that blue robots can only go some place purple robots have already visited. The purple robots are tasked with exploring the purple building while the blue robots are tasked with exploring the blue building.

enter a region after another set of vehicles has explored it. Both CLFs and CBFs can be combined with other objective functions that are task-oriented rather than safety oriented; these often then involve solving quadratic programs to satisfy safety constraints [275, 277, 279, 280]. The CLF/CBF approach is the most amenable to computation in high dimensional spaces, but in lower dimensional spaces one can directly solve for safety sets using reachability analysis, which depends on solving a Hamilton-Jacobi-Isaacs partial differential equation [281, 282]. Though not necessarily practical for high dimensional systems, this guarantees an optimal trade-off between safety and performance.

An important challenge in these safe learning techniques is that they are model-based. They require a model to evaluate the monotonic decrease of the CLFs/CBFs or to evaluate reachability conditions. Since a robot will typically be learning something about the environment relevant to its evolution, its own dynamics, or its interactions with the environment, all these techniques will rely on model updates of some form along with real-time updates to statistical analysis. A key question is how should a robot stay safe during this process, and what should safety mean when representations critical to safety are not known?

In recent work—following the CLF/CBF viewpoint of safety—we showed that one can use hybrid control methods to schedule switching between a safe controller and a learning controller, while maintaining the asymptotic properties of the safe controller [283]. The critical assumption in that work is that there is an operating point where stability of the robot-environment combination is already established and using the safety of that state as a starting point for safe learning. This is often a reasonable assumption; for instance, one might have an empirically safe PID controller for a humanoid robot near upright posture without having model-based safety analysis. Additionally, CBFs have been used to guide the learning process in reinforcement learning [284]. In this work, the CBFs restrict exploration to safe policies and become less conservative as an online learning process learns a model of the dynamical system. This makes the learning process more efficient while guaranteeing safety. This method incorporates online measurements to improve the CBF-RL controller, providing an opportunity for active learning approaches such as those discussed here to facilitate information gathering. Other approaches to simultaneously satisfying safety guarantees with *a priori* unknown dynamics and/or unknown environments need control formalisms that enforce safety criteria in the absence of any certainty.

7.3. Stability, Invariance, and Specification

Another concern critical to learning is how to impose prior knowledge on learned models. Particularly in the context of physical learning, where a model does not need to be an ordinary differential equation or a statistical pattern, but can instead be a principle (such as a motion symmetry [285] or energetic dissipation). Among these principled statements of modeling assumptions, stability, the property that the unforced system asymptotically converges to an equilibrium, may be the most common property in a physical system that we may wish to insist upon [286]. In [287]—following [81, 288–292]—we used recent results in linear algebra to project linear operators (such as the Koopman representations discussed earlier) onto the closest stable linear operators. Moreover, in [293] we applied these techniques to robotic manipulation examples, where notably the experiments were implausible without constraining the learning to stable models.

There is a wide range of potential specifications one may wish to impose on a learning system. How would one specify that a learned model must satisfy a linear temporal logic (LTL) constraint such as those described in [294]? What about symmetries in time and space, implying conservation of energy and momentum? Developing formally correct methods for combining learning tools with these specifications is a key step forward towards robot learning under user-generated constraints on what should be learned.

7.4. Actionable Learning

A key property of linear control systems is the separation principle. This principle asserts that an optimal estimator can be designed independently from the optimal control. A consequence of the separation principle is that as soon as a measurement has been taken, one knows that the automation system can start to productively take actions. That is, every measurement is *actionable* for the control system. A generalization of the separation principle is to ask whether designing a learning algorithm can be done independently from designing its control system. In general, this review assumes that this is not possible—the learning and control goals are mutually dependent. However, in some learning cases the relationship between what is being learned and when or how soon one can take action may be important. For instance, in the case of shape recognition in Figure 2, exploring an object to determine its shape properties must happen prior to exploring an unknown environment in search of that shape. This transition is an example of the representation (in this case the abstraction's "shape") becoming actionable to the control system. As far as the authors are aware, this topic is little studied in control, but has a long history in psychological study of decision making (*e.g.*, see the many books on this topic by Alain Berthoz [295]).

When a control system becomes actionable is particularly important when distinguishing between active learning and passive learning. During the active learning phase, learning may be the primary goal of the control system. During the passive learning phase the robotic system (or animal) may transition to attempting its ultimate task while continuing to run online passive learning updates. In single-shot learning, where the learner only has one trajectory to exploit for the purpose of learning, being able to robustly detect when learning has become sufficient to take action is a critical part of the path to task success.

Analysis methods are needed for describing conditions under which learned models are *sufficient* for making a decision to combine the estimation aspects of learning with the control aspects of learning. This transition is often characterized in terms of exploration/exploitation tradeoffs [296] in the context of sampling-based learning. In the context of a physical system, exploration and exploitation depend on the physics of the learner and environment, and the transition between them will be regulated by the control system. In the case of the example in Figure 3, this would be a safety-critical decision—devoting inadequate time for active learning yields an insufficient model for recovery prior to the vehicle hitting the ground, while engaging in active learning too long will lead to a catastrophic failure. This particular example would likely yield a convex function that represents safety as a function of transition time. However, how to analyze and compute this transition in general is unknown.

Efficiently forming representations relevant to task completion is part of the challenge in forming actionable representations. When a representation becomes actionable, we capture particular elements of the underlying object or task relevant for decision making while ignoring irrelevant sensory data. The question of determining functionally applicable representations has been explored in [297]. The authors claim that the structure of the environment can be modeled with a known goal-conditioned policy—a policy that can achieve a goal state from a given state. The authors refine this policy by differentiating states using the actions necessary to reach them. Thus, states that are functionally similar are closer to each other in the representation than they would be when representing their location with an Euclidean distance. This method could benefit from active learning. For instance, one may use the entropy of the representation rather than the entropy of the input or entropy of the physical states, as the information quantity to force active learning capabilities. However one constructs representations from data-driven experience, an important question will be how to synthesize active learning to close the loop on representation generation.

8. Conclusion

Active learning and data-driven control will play a major role in future robotic systems operating without access to reliable analytic models or prior data sets in uncertain environments. Robots will need to become fluent learners—routinely investing time and energy in single-shot learning through purposeful data collection and interpretation. This high level goal transcends the capabilities currently available for robotics in machine learning, both in terms of specifying behavior and representing learning goals. Machine intelligence in general has almost entirely been viewed as an extension of estimation theory, focusing on the processing of data. Even reinforcement learning assumes that the data needed for updating a policy is available or that it can be created in simulation. Here we view learning, in part, as an extension of control theory, focusing on how decisions impact learning outcomes. Before these two views can be synthesized into a single coherent theory, many challenges need to be addressed including those mentioned earlier and many not yet understood.

Expanding our notion of a *model* becomes a key effort moving forward. Models should no longer be solely defined by an ordinary differential equation, though ordinary differential equations may still play critical roles during analysis and computation. Instead, a theme in this review is that model-based reasoning needs to admit any set of meta-principles one asserts, such as symmetries in the system, its stability properties, what equilibria are expected, or its logical structure. These assertions will constrain numerical inference, thereby improving learning by reducing the classes of admissible models.

We have outlined and argued for the development of a theory of robot learning—one that deals with the difficulties and constraints that an embodied learning agent would face in the physical world. While much of machine learning has neglected the challenges that physical embodiment brings, this presents a great opportunity for control theorists at-large. The historical arc of robot control has retained a clear focus on the physical properties that ensure safe, robust, and reliable performance. By merging our understanding of controllability, stability, and compliance, with the flexibility of black-box learning, an action-oriented theory of learning will be key to enable future robot technologies.

Acknowledgements

We would like to thank Muchen Sun, Ana Pervan, Kyra Rudy, Frank Park, and the anonymous reviewers of the first draft for their many helpful comments on this manuscript.

References

[1] R. Bajcsy, Active perception, Proceedings of the IEEE 76 (8) (1988) 996–1005.

[2] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[3] S. C. Stanton, Situated experimental agents for scientific discovery, Science Robotics 3 (24) (2018) 4978.

[4] H. Martin, Osmotropotaxis in the honey-bee, Nature 208 (5005) (1965) 59–63.

[5] J. A. Basil, R. T. Hanlon, S. I. Sheikh, J. Atema, Three-dimensional odor tracking by nautilus pompilius, Journal of Experimental Biology 203 (9) (2000) 1409–1414.

[6] Y. Yovel, B. Falk, C. F. Moss, N. Ulanovsky, Optimal localization by pointing off axis, Science 327 (5966) (2010) 701–704.

[7] B. Webb, R. R. Harrison, M. A. Willis, Sensorimotor control of navigation in arthropod and artificial systems, Arthropod Structure and Development 33 (3) (2004) 301–329.

[8] A. G. Khan, M. Sarangi, U. S. Bhalla, Rats track odour trails accurately using a multi-layered strategy with near-optimal sampling, Nature Communications 3 (1) (2012) 1–10.

[9] S. A. Stamper, E. Roth, N. J. Cowan, E. S. Fortune, Active sensing via movement shapes spatiotemporal patterns of sensory feedback, Journal of Experimental Biology 215 (9) (2012) 1567–1574.

[10] K. C. Catania, Stereo and serial sniffing guide navigation to an odour source in a mammal, Nature Communications 4 (1) (2013) 1–8.

[11] M. J. Hartmann, Active sensing capabilities of the rat whisker system, Autonomous Robots 11 (3) (2001) 249–254.

[12] M. E. Nelson, M. A. MacIver, Sensory acquisition in active sensing systems, Journal of Comparative Physiology A 192 (6) (2006) 573–586.

[13] C. Chen, T. D. Murphey, M. A. MacIver, Tuning movement for sensing in an uncertain world, eLife 9 (2020) e52371.

[14] K. Nakajima, H. Hauser, T. Li, R. Pfeifer, Information processing via physical soft body, Scientific Reports 5 (1) (2015) 10487.

[15] X. Yin, R. Müller, Integration of deep learning and soft robotics for a biomimetic approach to nonlinear sensing, Nature Machine Intelligence (Apr 2021).

[16] T. Chen, M. Pauly, P. M. Reis, A reprogrammable mechanical metamaterial with stable memory, Nature 589 (7842) (2021) 386–390.

[17] J. M. Gold, J. L. England, Self-organized novelty detection in driven spin glasses (2019).

[18] W. Zhong, J. M. Gold, S. Marzen, J. L. England, N. Y. Halpern, Learning about learning by many-body systems (2020).

[19] R. Pfeifer, G. Gómez, Morphological computation–connecting brain, body, and environment, in: Creating brain-like intelligence, Springer, 2009, pp. 66–83.

[20] D. Rus, M. T. Tolley, Design, fabrication and control of soft robots, Nature 521 (7553) (2015) 467–475.

[21] N. Furutani, T. Takahashi, N. Naito, T. Maruishi, Y. Yoshimura, C. Hasegawa, T. Hirosawa, M. Kikuchi, Complexity of body movements during sleep in children with autism spectrum disorder, Entropy 23 (4) (2021).

[22] M. Osipov, Y. Behzadi, J. M. Kane, G. Petrides, G. D. Clifford, Objective identification and analysis of physiological and behavioral signs of schizophrenia, Journal of Mental Health 24 (5) (2015) 276–282.

[23] T. Berrueta, A. Pervan, K. Fitzsimons, T. Murphey, Dynamical system segmentation for information measures in motion, IEEE Robotics and Automation Letters 4 (1) (2019) 169–176.

[24] K. Fitzsimons, A. M. Acosta, J. Dewald, T. D. Murphey, Ergodicity reveals assistance and learning in physical human robot interaction, Science: Robotics 4 (29) (2019) 6079.

[25] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, M. G. E. da Luz, E. P. Raposo, H. E. Stanley, Optimizing the success of random searches, Nature 401 (6756) (1999) 911–914.

[26] F. Bartumeus, J. Catalan, Optimal search behavior and classic foraging theory, Journal of Physics A: Mathematical and Theoretical 42 (43) (2009) 434002.

[27] R. J. Baddeley, N. R. Franks, E. R. Hunt, Optimal foraging and the information theory of gambling, Journal of The Royal Society Interface 16 (157) (2019) 20190162.

[28] M. Bekoff, Animal Play: Problems and Perspectives, Springer US, 1976, pp. 165–188.

[29] A. S. Reinhold, J. I. Sanguinetti-Scheck, K. Hartmann, M. Brecht, Behavioral and neural correlates of hide-and-seek in rats, Science 365 (6458) (2019) 1180–1183.

[30] P. K. Smith, Does play matter? functional and evolutionary aspects of animal and human play, Behavioral and Brain Sciences 5 (1) (1982) 139–155.

[31] B. Settles, Active learning literature survey, Computer Sciences Technical Report 1648, University of Wisconsin–Madison (2009).

[32] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, T. Darrell, Deep learning for tactile understanding from visual and haptic data, in: 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016, pp. 536–543.

[33] C. Li, T. Zhang, D. I. Goldman, A terradynamics of legged locomotion on granular media, Science 339 (6126) (2013) 1408–1412.

[34] C. Laschi, B. Mazzolai, M. Cianchetti, Soft robotics: Technologies and systems pushing the boundaries of robot abilities, Science Robotics 1 (1) (2016).

[35] J.-P. Merlet, A historical perspective of robotics, in: M. Ceccarelli (Ed.), International Symposium on History of Machines and Mechanisms Proceedings HMM 2000, Springer Netherlands, Dordrecht, 2000, pp. 379–386.

[36] G. C. Devol, Programmable article transfer, U.S. Patent 2,988,237 (Dec 1954).

[37] W. Walter, A machine that learns, Scientific American 185 (1951) 60–63.

[38] A. M. Turing, On Computable Numbers, with an Application to the Entscheidungsproblem, Proceedings of the London Mathematical Society s2-42 (1) (1937) 230–265.

[39] A. M. Turing, Computing machinery and intelligence, Mind 59 (236) (1950) 433–460.

[40] J. McCarthy, P. J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: Readings in Artificial Intelligence, Elsevier, 1969, pp. 431–450.

[41] E. M. Gold, Language identification in the limit, Information and Control 10 (5) (1967) 447–474.

[42] D. Angluin, Inductive inference of formal languages from positive data, Information and Control 45 (2) (1980) 117–135.

[43] L. G. Valiant, A theory of the learnable, Communications of the ACM 27 (11) (1984) 1134–1142.

[44] F. Rosenblatt, The perceptron: A probabilistic model for information storage and organization in the brain., Psychological review 65 (6) (1958) 386–408.

[45] D. Angluin, C. H. Smith, Inductive inference: Theory and methods, ACM computing surveys (CSUR) 15 (3) (1983) 237–269.

[46] N. Littlestone, Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm, Machine Learning 2 (4) (1988) 285–318.

[47] M. Kearns, Efficient noise-tolerant learning from statistical queries, Journal of the ACM) 45 (6) (1998) 983–1006.

[48] S. Ben-David, A. Itai, E. Kushilevitz, Learning by distances, Information and Computation 117 (2) (1995) 240–250.

[49] S. Ben-David, P. Hrubeš, S. Moran, A. Shpilka, A. Yehudayoff, Learnability can be undecidable, Nature Machine Intelligence 1 (1) (2019) 44–48.

[50] A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth, Occam's razor, Information Processing Letters 24 (6) (1987) 377–380.

[51] A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth, Learnability and the Vapnik-Chervonenkis dimension, Journal of the ACM 36 (4) (1989) 929–965.

[52] S. B. Cooper, Computability Theory, CRC Press, 2003.

[53] D. Angluin, Inference of reversible languages, Journal of the ACM 29 (3) (1982) 741–765.

[54] L. Nocks, The Robot: The Life Story of a Technology, Greenwood Technographies, Greenwood Press, 2007.

[55] D. Angluin, Queries and concept learning, Machine Learning 2 (4) (1988) 319–342.

[56] D. A. Cohn, Z. Ghahramani, M. I. Jordan, Active learning with statistical models, Journal of Artificial Intelligence Research 4 (1996) 129–145.

[57] M.-F. F. Balcan, V. Feldman, Statistical active learning algorithms, in: Advances in Neural Information Processing Systems (NeurIPS), Vol. 26, 2013.

[58] M. Balcan, S. Hanneke, J. W. Vaughan, The true sample complexity of active learning, Machine Learning 80 (2) (2010) 111–139.

[59] J. B. Watson, Psychology as the behaviorist views it, Psychological Review 20 (2) (1913) 158–177.

[60] B. F. Skinner, The Behavior of Organisms: An Experimental Analysis, Appleton-Century-Crofts, 1938.

[61] A. G. Barto, R. S. Sutton, P. S. Brouwer, Associative search network: A reinforcement learning associative memory, Biological cybernetics 40 (3) (1981) 201–211.

[62] R. S. Sutton, A. G. Barto, Toward a modern theory of adaptive networks: Expectation and prediction, Psychological Review 88 (2) (1981) 135–170.

[63] A. G. Barto, R. S. Sutton, C. W. Anderson, Neuronlike adaptive elements that can solve difficult learning control problems, IEEE Transactions on Systems, Man, and Cybernetics SMC-13 (5) (1983) 834–846.

[64] R. Bellman, Dynamic programming, Science 153 (3731) (1966) 34–37.

[65] R. S. Sutton, A. G. Barto, R. J. Williams, Reinforcement learning is direct adaptive optimal control, IEEE Control Systems Magazine 12 (2) (1992) 19–22.

[66] E. L. Thorndike, The law of effect, The American Journal of Psychology 39 (1) (1927) 212–222.

[67] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[68] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning., in: Proceedings of the International Conference on Learning Representations (ICLR), 2016.

[69] Y. Duan, X. Chen, R. Houthooft, J. Schulman, P. Abbeel, Benchmarking deep reinforcement learning for continuous control, in: Proceedings of the International Conference on Machine Learning (ICML), Vol. 48, 2016, pp. 1329–1338.

[70] S. Gu, E. Holly, T. Lillicrap, S. Levine, Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2017, pp. 3389–3396.

[71] L. P. Kaelbling, The foundation of efficient robot learning, Science 369 (6506) (2020) 915–916.

[72] J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, S. Levine, How to train your robot with deep reinforcement learning: lessons we have learned, The International Journal of Robotics Research (2021).

[73] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, P. Corke, The limits and potentials of deep learning for robotics, The International Journal of Robotics Research 37 (4-5) (2018) 405–420.

[74] T. Haarnoja, H. Tang, P. Abbeel, S. Levine, Reinforcement learning with deep energy-based policies, in: Proceedings of the International Conference on Machine Learning (ICML), Vol. 70, 2017, pp. 1352–1361.

[75] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, in: Proceedings of the International Conference on Machine Learning (ICML), Vol. 80, 2018, pp. 1861–1870.

[76] B. Eysenbach, S. Levine, Maximum entropy RL (provably) solves some robust RL problems (2021).

[77] X. B. Peng, M. Andrychowicz, W. Zaremba, P. Abbeel, Sim-to-real transfer of robotic control with dynamics randomization, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 3803–3810.

[78] A. A. Rusu, M. Večerík, T. Rothörl, N. Heess, R. Pascanu, R. Hadsell, Sim-to-real robot learning from pixels with progressive nets, in: Proceedings of the 1st Annual Conference on Robot Learning, Vol. 78 of Proceedings of Machine Learning Research, PMLR, 2017, pp. 262–270.

[79] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, K. Bousmalis, Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[80] J. García, F. Fernández, A comprehensive survey on safe reinforcement learning, Journal of Machine Learning Research 16 (42) (2015) 1437–1480.

[81] J. Z. Kolter, G. Manek, Learning stable deep dynamics models, in: Advances in Neural Information Processing Systems, 2019, pp. 11126–11134.

[82] C. Gehring, D. Precup, Smart exploration in reinforcement learning using absolute temporal difference errors, in: Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems, 2013, pp. 1037–1044.

[83] A. Tsiamis, G. J. Pappas, Linear systems can be hard to learn (2021).

[84] A. K. Akametalu, J. F. Fisac, J. H. Gillula, S. Kaynama, M. N. Zeilinger, C. J. Tomlin, Reachability-based safe learning with Gaussian processes, in: 53rd IEEE Conference on Decision and Control (CDC), 2014, pp. 1424–1431.

[85] H. J. S. Feder, J. J. Leonard, C. M. Smith, Adaptive mobile robot navigation and mapping, International Journal of Robotics Research 18 (7) (1999) 650–668.

[86] C. Leung, S. Huang, N. Kwok, G. Dissanayake, Planning under uncertainty using model predictive control for information gathering, Robotics and Autonomous Systems 54 (11) (2006) 898–910.

[87] R. Sim, N. Roy, Global A-optimal robot exploration in SLAM, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2005, pp. 661–666.

[88] J. Vander Hook, P. Tokekar, V. Isler, Cautious greedy strategy for bearing-based active localization: Experiments and theoretical analysis, in: IEEE International Conference on Robotics and Automation (ICRA), 2012, pp. 1787–1792.

[89] R. Marchant, F. Ramos, Bayesian optimisation for intelligent environmental monitoring, in: IEEE Int. Conf. on Intelligent Robots and Systems (IROS), 2012, pp. 2242–2249.

[90] E.-M. Wong, F. Bourgault, T. Furukawa, Multi-vehicle Bayesian search for multiple lost targets, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2005, pp. 3169–3174.

[91] C. Stachniss, W. Burgard, Exploring unknown environments with mobile robots using coverage maps, in: International Joint Conference on Artificial Intelligence, 2003, pp. 1127–1134.

[92] C. Kreucher, J. Wegrzyn, M. Beauvais, R. Conti, Multi-

platform information-based sensor management: an inverted UAV demonstration, in: SPIE Defense Transformation and Network-Centric Systems, Vol. 6578, 2007, pp. 65780Y–1–65780Y–11.

[93] N. Roy, C. Earnest, Dynamic action spaces for information gain maximization in search and exploration, in: American Controls Conf. (ACC), 2006, pp. 1631–1636.

[94] W. Lu, G. Zhang, S. Ferrari, R. Fierro, I. Palunko, An information potential approach for tracking and surveilling multiple moving targets using mobile sensor agents, in: SPIE Unmanned Systems Technology, Vol. 8045, 2011, pp. 80450T–1–80450T–13.

[95] F. Bourgault, A. A. Makarenko, S. Williams, B. Grocholsky, H. Durrant-Whyte, Information based adaptive robotic exploration, in: IEEE Int. Conf. on Intelligent Robots and Systems (IROS), Vol. 1, 2002, pp. 540 – 545.

[96] A. Elfes, Using occupancy grids for mobile robot perception and navigation, Computer 22 (6) (1989) 46 –57.

[97] A. Bender, S. B. Williams, O. Pizarro, Autonomous exploration of large-scale benthic environments, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2013, pp. 390–396.

[98] N. Cao, K. H. Low, J. M. Dolan, Multi-robot informative path planning for active sensing of environmental phenomena: A tale of two algorithms, in: International Conference on Autonomous Agents and Multi-agent Systems, 2013, pp. 7–14.

[99] T. N. Hoang, K. H. Low, P. Jaillet, M. Kankanhalli, Nonmyopic ϵ-Bayes-optimal active learning of Gaussian processes, in: International Conference on Machine Learning, 2014, pp. 739–747.

[100] K. H. Low, J. M. Dolan, P. Khosla, Adaptive multi-robot wide-area exploration and mapping, in: Conference on Autonomous Agents and Multiagent Systems, 2008, pp. 23–30.

[101] A. Singh, A. Krause, C. Guestrin, W. J. Kaiser, Efficient informative sensing using multiple robots, Journal of Artificial Intelligence Research (JAIR) 34 (2009) 707–755.

[102] J. Souza, R. Marchant, L. Ott, D. Wolf, F. Ramos, Bayesian optimisation for active perception and smooth navigation, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2014, pp. 4081–4087.

[103] G. Picardi, M. Chellapurath, S. Iacoponi, S. Stefanni, C. Laschi, M. Calisti, Bioinspired underwater legged robot for seabed exploration with low environmental disturbance, Science Robotics 5 (42) (2020).

[104] J. A. Breier, M. V. Jakuba, M. A. Saito, G. J. Dick, S. L. Grim, E. W. Chan, M. R. McIlvin, D. M. Moran, B. A. Alanis, A. E. Allen, C. L. Dupont, R. Johnson, Revealing ocean-scale biochemical structure with a deep-diving vertical profiling autonomous vehicle, Science Robotics 5 (48) (2020).

[105] Y. Zhang, J. P. Ryan, B. W. Hobson, B. Kieft, A. Romano, B. Barone, C. M. Preston, B. Roman, B.-Y. Raanan, D. Pargett, M. Dugenne, A. E. White, F. H. Freitas, S. Poulos, S. T. Wilson, E. F. DeLong, D. M. Karl, J. M. Birch, J. G. Bellingham, C. A. Scholin, A system of coordinated autonomous robots for lagrangian studies of microbes in the oceanic deep chlorophyll maximum, Science Robotics 6 (50) (2021).

[106] T. O. Fossum, G. M. Fragoso, E. J. Davies, J. E. Ullgren, R. Mendes, G. Johnsen, I. Ellingsen, J. Eidsvik, M. Ludvigsen, K. Rajan, Toward adaptive robotic sampling of phytoplankton in the coastal ocean, Science Robotics 4 (27) (2019).

[107] I. Abraham, A. Mavrommati, T. D. Murphey, Data-driven measurement models for active localization in sparse environments, in: Robotics: Science and Systems Proceedings, 2018.

[108] A. Guntuboyina, B. Sen, et al., Nonparametric shape-restricted regression, Statistical Science 33 (4) (2018) 568–594.

[109] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, H.-P. Seidel, Multilinear pose and body shape estimation of dressed subjects from image sets, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 1823–1830.

[110] B. Southall, C. Taylor, Stochastic road shape estimation, in: Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV), Vol. 1, 2001, pp. 205–212.

[111] I. Abraham, A. Prabhakar, M. Hartmann, T. Murphey, Ergodic exploration using binary sensing for non-parametric shape estimation, IEEE Robotics and Automation Letters 2 (2) (2017) 827–834.

[112] A. M. Bayen, I. M. Mitchell, M. M. Oishi, C. J. Tomlin, Aircraft autolander safety analysis through optimal control-based reach set computation, Journal of Guidance, Control, and Dynamics 30 (1) (2007) 68–77.

[113] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (5923) (2009) 81–85.

[114] I. Abraham, T. D. Murphey, Active learning of dynamics for data-driven control using Koopman operators, IEEE Transactions on Robotics 35 (5) (2019) 1071–1083.

[115] T. A. Berrueta, I. Abraham, T. Murphey, Experimental Applications of the Koopman Operator in Active Learning for Control, Springer International Publishing, 2020, pp. 421–450.

[116] M. Oubbati, G. Palm, A neural framework for adaptive robot control, Neural Computing and Applications 19 (1) (2010) 103–114.

[117] D. Nguyen-Tuong, M. Seeger, J. Peters, Model learning with local Gaussian process regression, Advanced Robotics 23 (15) (2009) 2015–2034.

[118] J. Z. Kim, Z. Lu, E. Nozari, G. J. Pappas, D. S. Bassett, Teaching recurrent neural networks to infer global temporal structure from local examples, Nature Machine Intelligence 3 (4) (2021) 316–323.

[119] P. Karkus, X. Ma, D. Hsu, L. P. Kaelbling, W. S. Lee, T. Lozano-Pérez, Differentiable algorithm networks for composable robot learning, Robotics: Science and Systems (2019).

[120] T. G. Thuruthel, B. Shih, C. Laschi, M. T. Tolley, Soft robot perception using embedded soft sensors and recurrent neural networks, Science Robotics 4 (26) (2019).

[121] T. Hofmann, B. Schölkopf, A. J. Smola, Kernel methods in machine learning, The Annals of Statistics (2008) 1171–1220.

[122] S. Schaal, C. Atkeson, S. Vijayakumar, Real-time robot learning with locally weighted statistical learning, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Vol. 1, 2000, pp. 288–293.

[123] C.-A. Cheng, H.-P. Huang, H.-K. Hsu, W.-Z. Lai, C.-C. Cheng, Learning the inverse dynamics of robotic manipulators in structured reproducing kernel Hilbert space, IEEE Transactions on Cybernetics 46 (7) (2016) 1691–1703.

[124] A. Dalla Libera, R. Carli, A data-efficient geometrically inspired polynomial kernel for robot inverse dynamic, IEEE Robotics and Automation Letters 5 (1) (2019) 24–31.

[125] A. J. Smola, B. Schölkopf, Bayesian kernel methods, in: Advanced lectures on machine learning, Springer, 2003, pp. 65–117.

[126] R. G. Gallager, Stochastic Processes: Theory for Applications, Cambridge University Press, 2013.

[127] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2005.

[128] M. P. Deisenroth, D. Fox, C. E. Rasmussen, Gaussian processes for data-efficient learning in robotics and control, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (2) (2015) 408–423.

[129] S. E. Otto, C. W. Rowley, Koopman operators for estimation and control of dynamical systems, Annual Review of Control, Robotics, and Autonomous Systems 4 (1) (2021).

[130] B. Koopman, Hamiltonian systems and transformation in Hilbert space, Proc. National Academy of Sciences 17 (5) (1931) 315–318.

[131] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, J. N. Kutz, On dynamic mode decomposition: Theory and applications, J. Computational Dynamics 1 (2014) 391.

[132] M. O. Williams, I. G. Kevrekidis, C. W. Rowley, A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition, J. Nonlinear Science 25 (2015) 1307–1346.

[133] S. Brunton, B. Brunton, J. Proctor, J. Kutz, Koopman in-

variant subspaces and finite linear representations of nonlinear dynamical systems for control, PLOS ONE 11 (2) (2016) 1–19.

[134] J. L. Proctor, S. L. Brunton, J. N. Kutz, Generalizing Koopman theory to allow for inputs and control, SIAM Journal on Applied Dynamical Systems 17 (1) (2018) 909–930.

[135] E. Kaiser, J. N. Kutz, S. L. Brunton, Data-driven discovery of Koopman eigenfunctions for control, in: arXiv, 2017.

[136] I. Abraham, G. de la Torre, T. Murphey, Model-based control using Koopman operators, in: Robotics: Science and Systems Proceedings, 2017.

[137] D. Bruder, B. Gillespie, C. D. Remy, R. Vasudevan, Modeling and control of soft robots using the Koopman operator and model predictive control, in: Robotics: Science and Systems, 2019.

[138] G. Mamakoukas, M. L. Castaño, X. Tan, T. D. Murphey, Local Koopman operators for data-driven control of robotic systems, in: Robotics: Science and Systems, 2019.

[139] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006.

[140] G. Zhong, L.-N. Wang, X. Ling, J. Dong, An overview on data representation learning: From traditional feature learning to recent deep learning, The Journal of Finance and Data Science 2 (4) (2016) 265–278.

[141] J. Butepage, M. J. Black, D. Kragic, H. Kjellstrom, Deep representation learning for human motion prediction and classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6158–6166.

[142] J. Liu, Z. Liu, L. Wang, L. Guo, J. Dang, Time-frequency deep representation learning for speech emotion recognition integrating self-attention, in: T. Gedeon, K. W. Wong, M. Lee (Eds.), Neural Information Processing, Springer International Publishing, 2019, pp. 681–689.

[143] T. de Bruin, J. Kober, K. Tuyls, R. Babuška, Integrating state representation learning into deep reinforcement learning, IEEE Robotics and Automation Letters 3 (3) (2018) 1394–1401.

[144] B. Lusch, J. N. Kutz, S. L. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, Nature Communications 9 (1) (2018) 4950.

[145] B. Shih, D. Shah, J. Li, T. G. Thuruthel, Y.-L. Park, F. Iida, Z. Bao, R. Kramer-Bottiglio, M. T. Tolley, Electronic skins and machine learning for intelligent soft robots, Science Robotics 5 (41) (2020).

[146] H. Madokoro, K. Sato, N. Shimoi, Adaptive category mapping networks for all-mode topological feature learning used for mobile robot vision, in: Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE, 2014, pp. 678–683.

[147] J. R. Spletzer, C. J. Taylor, Dynamic sensor planning and control for optimally tracking targets, International Journal of Robotics Research 22 (1) (2003) 7–20.

[148] B. DasGupta, J. P. Hespanha, J. Riehl, E. Sontag, Honey-pot constrained searching with local sensory information, Nonlinear Analysis: Theory, Methods & Applications 65 (9) (2006) 1773–1793.

[149] G. Zhang, S. Ferrari, An adaptive artificial potential function approach for geometric sensing, in: IEEE Int. Conf. on Decision and Control (CDC), 2009, pp. 7903–7910.

[150] G. Hager, M. Mintz, Computational methods for task-directed sensor data fusion and sensor planning, International Journal of Robotics Research 10 (4) (1991) 285–313.

[151] G. Benet, F. Blanes, J. Simó, P. Pérez, Using infrared sensors for distance measurement in mobile robots, Robotics and Autonomous Systems 40 (4) (2002) 255 – 266.

[152] J. Denzler, M. Zobel, H. Niemann, Information theoretic focal length selection for real-time active 3d object tracking, in: IEEE Int. Conf. on Computer Vision, 2003, pp. 400–407.

[153] M. Vergassola, E. Villermaux, B. I. Shraiman, Infotaxis as a strategy for searching without gradients, Nature 445 (7126) (2007) 406.

[154] D. Fox, W. Burgard, S. Thrun, Active Markov localization for mobile robots, Robotics and Autonomous Systems 25 (3-4) (1998) 195–207.

[155] T. Arbel, F. Ferrie, Viewpoint selection by navigation through entropy maps, in: IEEE Int. Conf. on Computer Vision, 1999, pp. 248–254.

[156] P.-P. Vázquez, M. Feixas, M. Sbert, W. Heidrich, Viewpoint selection using viewpoint entropy., in: Vision Modeling and Visualization Conference, Vol. 1, 2001, pp. 273–280.

[157] Y. Takeuchi, N. Ohnishi, N. Sugie, Active vision system based on information theory, Systems and Computers in Japan 29 (11) (1998) 31–39.

[158] C. Kreucher, K. Kastella, A. O. Hero, Sensor management using an active sensing approach, Signal Processing 85 (3) (2005) 607–624.

[159] J. Toh, S. Sukkarieh, A Bayesian formulation for the prioritized search of moving objects, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2006, pp. 219–224.

[160] J. Denzler, C. Brown, Information theoretic sensor data selection for active object recognition and state estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2) (2002) 145–157.

[161] J. Tisdale, Z. Kim, J. K. Hedrick, Autonomous UAV path planning and estimation, IEEE Robotics and Automation Magazine 16 (2) (2009) 35–42.

[162] B. Grocholsky, J. Keller, V. Kumar, G. Pappas, Cooperative air and ground surveillance, IEEE Robotics and Automation Magazine 13 (3) (2006) 16–25.

[163] W. Lu, G. Zhang, S. Ferrari, An information potential approach to integrated sensor path planning and control, IEEE Transactions on Robotics 30 (4) (2014) 919–934.

[164] G. Zhang, S. Ferrari, M. Qian, An information roadmap method for robotic sensor path planning, Journal of Intelligent and Robotic Systems 56 (1-2) (2009) 69–98.

[165] G. A. Hollinger, B. Englot, F. S. Hover, U. Mitra, G. S. Sukhatme, Active planning for underwater inspection and the benefit of adaptivity, International Journal of Robotics Research 32 (1) (2013) 3–18.

[166] X. Liao, L. Carin, Application of the theory of optimal experiments to adaptive electromagnetic-induction sensing of buried targets, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (8) (2004) 961–972.

[167] A. Emery, A. V. Nenarokomov, Optimal experiment design, Measurement Science and Technology 9 (6) (1998) 864.

[168] D. Ucinski, J. Korbicz, Path planning for moving sensors in parameter estimation of distributed systems, in: Workshop on Robot Motion and Control (RoMoCo), 1999, pp. 273–278.

[169] D. Ucinski, Optimal sensor location for parameter estimation of distributed processes, International Journal of Control 73 (13) (2000) 1235–1248.

[170] R. B. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, 2004.

[171] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 379–423.

[172] N. Atanasov, B. Sankaran, J. Le Ny, G. Pappas, K. Daniilidis, Nonmyopic view planning for active object classification and pose estimation, IEEE Transactions on Robotics 30 (5) (2014) 1078–1090.

[173] Y. F. Li, Z. G. Liu, Information entropy based viewpoint planning for 3-D object reconstruction, IEEE Transactions on Robotics 21 (3) (2005) 324–327.

[174] M. Rahimi, M. Hansen, W. Kaiser, G. Sukhatme, D. Estrin, Adaptive sampling for environmental field estimation using robotic sensors, in: IEEE Int. Conf. on Intelligent Robots and Systems (IROS), 2005, pp. 3692–3698.

[175] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley & Sons, 2012.

[176] N. Wahlström, T. B. Schön, M. P. Deisenroth, From pixels to torques: Policy learning with deep dynamical models (2015).

[177] B. Tovar, L. Munoz-Gomez, R. Murrieta-Cid, M. Alencastre-Miranda, R. Monroy, S. Hutchinson, Planning exploration strategies for simultaneous localization and mapping, Robotics and Autonomous Systems 54 (4) (2006) 314 – 331.

[178] B. Tovar, T. D. Murphey, Trajectory tracking among landmarks and binary sensor beams, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2012, pp. 2121–2127.

[179] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, J. J. Leonard, Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age, IEEE Transactions on Robotics 32 (6) (2016) 1309–1332.

[180] J. v. Neumann, Proof of the quasi-ergodic hypothesis, Proceedings of the National Academy of Sciences 18 (1) (1932) 70–82.

[181] U. Krengel, Ergodic theorems, de Gruyter, 1985.

[182] D. A. Shell, C. V. Jones, M. J. Matarić, Ergodic dynamics by design: A route to predictable multi-robot systems, in: Multi-Robot Systems. From Swarms to Intelligent Automata, Springer, 2005, pp. 291–297.

[183] G. Mathew, I. Mezić, Metrics for ergodicity and design of ergodic dynamics for multi-agent systems, Physica D: Nonlinear Phenomena 240 (4) (2011) 432–442.

[184] A. Wilson, J. Schultz, T. D. Murphey, Trajectory synthesis for Fisher information maximization, IEEE Transactions on Robotics 30 (6) (2014) 1358–1370.

[185] J. Cooper, M. Goodrich, Towards combining UAV and sensor operator roles in UAV-enabled visual search, in: IEEE Int. Conf. on Human Robot Interaction (HRI), 2008, pp. 351–358.

[186] C. Cai, S. Ferrari, Information-driven sensor path planning by approximate cell decomposition, IEEE Transactions on Systems, Man, and Cybernetics 39 (3) (2009) 672–689.

[187] Y. Ye, J. K. Tsotsos, Sensor planning for 3D object search, Computer Vision and Image Understanding 73 (2) (1999) 145 – 168.

[188] N. A. Massios, R. B. Fisher, A best next view selection algorithm incorporating a quality criterion, in: British Machine Vision Conference, 1998, pp. 78.1–78.10.

[189] R. Marchant, F. Ramos, Bayesian optimisation for informative continuous path planning, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2014, pp. 6136–6143.

[190] D. Mayne, A second-order gradient method for determining optimal trajectories of non-linear discrete-time systems, International Journal of Control 3 (1) (1966) 85–95.

[191] R. Bellman, On the theory of dynamic programming, in: Proceedings of the National Academy, Vol. 38.8, 1952, p. 716.

[192] Y. Tassa, N. Mansard, E. Todorov, Control-limited differential dynamic programming, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2014, pp. 1168–1175.

[193] V. Kumar, A. Gupta, E. Todorov, S. Levine, Learning dexterous manipulation policies from experience and imitation, arXiv (2016).

[194] G. Williams, N. Wagener, B. Goldfain, P. Drews, J. M. Rehg, B. Boots, E. A. Theodorou, Information theoretic MPC for model-based reinforcement learning, International Conference on Robotics and Automation (ICRA) (2017).

[195] E. A. Theodorou, E. Todorov, Relative entropy and free energy dualities: Connections to path integral and KL control, in: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), 2012, pp. 1466–1473.

[196] H. J. Kappen, Path integrals and symmetry breaking for optimal control theory, Journal of Statistical Mechanics: Theory and Experiment 2005 (11) (2005) P11011–P11011.

[197] G. Williams, A. Aldrich, E. A. Theodorou, Model predictive path integral control: From theory to parallel computation, Journal of Guidance, Control, and Dynamics 40 (2) (2017) 344–357.

[198] A. Ansari, T. D. Murphey, Sequential action control: Closed-form optimal control for nonlinear and nonsmooth systems, IEEE Trans. Rob. 32 (2017).

[199] I. Abraham, A. Broad, A. Pinosky, B. Argall, T. D. Murphey, Hybrid control for learning motor skills, in: Workshop on the Algorithmic Foundations of Robotics (WAFR), 2020.

[200] A. Wilson, J. Schultz, A. Ansari, T. D. Murphey, Real-time trajectory synthesis for information maximization using Sequential Action Control and least-squares estimation, in: IEEE Int. Conf. on Intelligent Robots and Systems (IROS), 2015, pp. 4935–4940.

[201] G. A. Hollinger, G. S. Sukhatme, Sampling-based robotic information gathering algorithms, International Journal of Robotics Research 33 (9) (2014) 1271–1287.

[202] A. Ryan, J. K. Hedrick, Particle filter based information-theoretic active sensing, Robotics and Autonomous Systems 58 (5) (2010) 574 – 584.

[203] N. J. Vickers, Mechanisms of animal navigation in odor plumes, The Biological Bulletin 198 (2) (2000) 203–212.

[204] L. Miller, Y. Silverman, M. A. MacIver, T. Murphey, Ergodic exploration of distributed information, IEEE Transactions on Robotics 32 (1) (2016) 36–52.

[205] H. Nishimura, M. Schwager, SACBP: belief space planning for continuous-time dynamical systems via stochastic sequential action control, in: Workshop on the Algorithmic Foundations of Robotics, 2018, pp. 267–283.

[206] L. Dressel, M. J. Kochenderfer, Tutorial on the generation of ergodic trajectories with projection-based gradient descent, IET Cyber-Physical Systems: Theory & Applications 4 (2) (2019) 89–100.

[207] L. Dressel, M. J. Kochenderfer, Using neural networks to generate information maps for mobile sensors, in: 2018 IEEE Conference on Decision and Control (CDC), 2018, pp. 2555–2560.

[208] D. A. Paley, A. Wolek, Mobile sensor networks and control: Adaptive sampling of spatiotemporal processes, Annual Review of Control, Robotics, and Autonomous Systems 3 (2020) 91–114.

[209] Z. Chen, L. Xiao, Q. Wang, Z. Wang, Z. Sun, Coverage control of multi-agent systems for ergodic exploration, in: 2020 39th Chinese Control Conference (CCC), 2020, pp. 4947–4952.

[210] R. Khodayi-mehr, W. Aquino, M. M. Zavlanos, Model-based active source identification in complex environments, IEEE Transactions on Robotics 35 (3) (2019) 633–652.

[211] C. Veitch, D. Render, A. Aravind, Ergodic flocking, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2019, pp. 6957–6962.

[212] H. Salman, E. Ayvali, H. Choset, Multi-agent ergodic coverage with obstacle avoidance, in: International Conference on Automated Planning and Scheduling, 2017.

[213] E. Ayvali, H. Salman, H. Choset, Ergodic coverage in constrained environments using stochastic trajectory optimization, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 5204–5210.

[214] A. Prabhakar, I. Abraham, M. Schlafly, A. Taylor, K. Popovic, G. Diniz, B. Teich, B. Simidchieva, S. Clark, T. Murphey, Ergodic specifications for flexible swarm control: From user commands to persistent adaptation, in: Robotics: Science and Systems Proceedings, 2020.

[215] G. De La Torre, K. Flaßkamp, A. Prabhakar, T. D. Murphey, Ergodic exploration with stochastic sensor dynamics, in: American Controls Conf. (ACC), 2016, pp. 2971 – 2976.

[216] A. Mavrommati, E. Tzorakoleftherakis, I. Abraham, T. D. Murphey, Real-time area coverage and target localization using receding-horizon ergodic exploration, IEEE Transactions on Robotics 34 (1) (2018) 62–80.

[217] L. M. Miller, T. D. Murphey, Trajectory optimization for continuous ergodic exploration, in: American Controls Conf. (ACC), 2013, pp. 4196–4201.

[218] L. Miller, T. D. Murphey, Optimal planning for target localization and coverage using range sensing, in: IEEE Int. Conf. on Automation Science and Engineering (CASE), 2015, pp. 501–508.

[219] N. Agharese, T. Cloyd, L. H. Blumenschein, M. Raitor, E. W. Hawkes, H. Culbertson, A. M. Okamura, Hapwrap: Soft growing wearable haptic device, in: IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 5466–5472.

[220] K. C. Galloway, K. P. Becker, B. Phillips, J. Kirby, S. Licht, D. Tchernov, R. J. Wood, D. F. Gruber, Soft robotic grippers

for biological sampling on deep reefs, Soft Robotics 3 (1) (2016) 23–33.

[221] M. T. Tolley, R. F. Shepherd, B. Mosadegh, K. C. Galloway, M. Wehner, M. Karpelson, R. J. Wood, G. M. Whitesides, A resilient, untethered soft robot, Soft Robotics 1 (3) (2014) 213–223.

[222] M. T. Gillespie, C. M. Best, E. C. Townsend, D. Wingate, M. D. Killpack, Learning nonlinear dynamic models of soft robots for model predictive control with neural networks, in: 2018 IEEE International Conference on Soft Robotics (RoboSoft), 2018, pp. 39–45.

[223] C. Laschi, M. Cianchetti, Soft robotics: new perspectives for robot bodyware and control, Frontiers in bioengineering and biotechnology 2 (2014) 3.

[224] G. Picardi, H. Hauser, C. Laschi, M. Calisti, Morphologically induced stability on an underwater legged robot with a deformable body, International Journal of Robotics Research 40 (2019) 435–448.

[225] F. Mammano, R. Nobili, Biophysics of the cochlea: linear approximation, The Journal of the Acoustical Society of America 93 (6) (1993) 3320–3332.

[226] G. Sumbre, G. Fiorito, T. Flash, B. Hochner, Motor control of flexible octopus arms, Nature 433 (7026) (2005) 595–596.

[227] N. Sornkarn, T. Nanayakkara, Can a soft robotic probe use stiffness control like a human finger to improve efficacy of haptic perception?, IEEE Transactions on Haptics 10 (2) (2017) 183–195.

[228] R. R. Murphy, S. Tadokoro, A. Kleiner, Disaster robotics, in: Springer Handbook of Robotics, Springer, 2016, pp. 1577–1604.

[229] L. Paull, S. Saeedi, M. Seto, H. Li, Sensor-driven online coverage planning for autonomous underwater vehicles, IEEE/ASME Transactions on Mechatronics 18 (6) (2013) 1827–1838.

[230] T. Oksanen, A. Visala, Coverage path planning algorithms for agricultural field machines, Journal of Field Robotics 26 (8) (2009) 651–668.

[231] B. Englot, F. S. Hover, Sampling-based coverage path planning for inspection of complex structures, in: International Conference on Automated Planning and Scheduling (ICAPS), 2012.

[232] E. Galceran, M. Carreras, A survey on coverage path planning for robotics, Robotics and Autonomous Systems 61 (12) (2013) 1258–1276.

[233] H. Choset, Coverage for robotics–a survey of recent results, Annals of Mathematics and Artificial Intelligence 31 (1-4) (2001) 113–126.

[234] N. Karapetyan, K. Benson, C. McKinney, P. Taslakian, I. Rekleitis, Efficient multi-robot coverage of a known environment, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 1846–1852.

[235] G. E. Jan, C. Luo, L. Hung, S. Shih, A computationally efficient complete area coverage algorithm for intelligent mobile robot navigation, in: 2014 International Joint Conference on Neural Networks (IJCNN), 2014, pp. 961–966.

[236] M. Schwager, D. Rus, J.-J. Slotine, Decentralized, adaptive coverage control for networked robots, The International Journal of Robotics Research 28 (3) (2009) 357–375.

[237] Y. Stergiopoulos, A. Tzes, Spatially distributed area coverage optimisation in mobile robotic networks with arbitrary convex anisotropic patterns, Automatica 49 (1) (2013) 232–237.

[238] V. García-Garrido, A. Mancho, S. Wiggins, C. Mendoza, A dynamical systems approach to the surface search for debris associated with the disappearance of flight MH370, Nonlin. Processes Geophys 22 (2015) 701–712.

[239] M. L. Rodríguez-Arévalo, J. Neira, J. A. Castellanos, On the importance of uncertainty representation in active SLAM, IEEE Transactions on Robotics 34 (3) (2018) 829–834.

[240] C. Stachniss, G. Grisetti, W. Burgard, Information gain-based exploration using Rao-Blackwellized particle filters., in: Robotics: Science and Systems, Vol. 2, 2005, pp. 65–72.

[241] H. Carrillo, P. Dames, V. Kumar, J. A. Castellanos, Autonomous robotic exploration using occupancy grid maps and graph SLAM based on Shannon and Rényi entropy, in: IEEE international conference on robotics and automation (ICRA), 2015, pp. 487–494.

[242] L. Carlone, J. Du, M. K. Ng, B. Bona, M. Indri, Active SLAM and exploration with particle filters using Kullback-Leibler divergence, Journal of Intelligent & Robotic Systems 75 (2) (2014) 291–311.

[243] C. Leung, S. Huang, G. Dissanayake, Active SLAM using model predictive control and attractor based exploration, in: 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 5026–5031.

[244] N. Atanasov, J. Le Ny, K. Daniilidis, G. J. Pappas, Decentralized active information acquisition: Theory and application to multi-robot SLAM, in: 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 4775–4782.

[245] B. Bonet, H. Geffner, Planning with incomplete information as heuristic search in belief space, in: Proceedings of the Fifth International Conference on Artificial Intelligence Planning Systems, AAAI Press, 2000, p. 52–61.

[246] R. Platt Jr, R. Tedrake, L. Kaelbling, T. Lozano-Perez, Belief space planning assuming maximum likelihood observations, in: Robotics: Science and Systems, 2006.

[247] S. Prentice, N. Roy, The belief roadmap: Efficient planning in belief space by factoring the covariance, The International Journal of Robotics Research 28 (11-12) (2009) 1448–1465.

[248] R. Valencia, M. Morta, J. Andrade-Cetto, J. M. Porta, Planning reliable paths with pose SLAM, IEEE Transactions on Robotics 29 (4) (2013) 1050–1059.

[249] S. Patil, G. Kahn, M. Laskey, J. Schulman, K. Goldberg, P. Abbeel, Scaling up Gaussian belief space planning through covariance-free trajectory optimization and automatic differentiation, in: Workshop on the Algorithmic Foundations of Robotics, Springer, 2015, pp. 515–533.

[250] A. Kim, R. M. Eustice, Active visual SLAM for robotic area coverage: Theory and experiment, The International Journal of Robotics Research 34 (4-5) (2015) 457–475.

[251] T. Taketomi, H. Uchiyama, S. Ikeda, Visual SLAM algorithms: a survey from 2010 to 2016, IPSJ Transactions on Computer Vision and Applications 9 (1) (2017) 16.

[252] S. Chen, Y. Li, N. M. Kwok, Active vision in robotic systems: A survey of recent developments, The International Journal of Robotics Research 30 (11) (2011) 1343–1377.

[253] V. Mnih, N. Heess, A. Graves, et al., Recurrent models of visual attention, Advances in Neural Information Processing Systems 27 (2014) 2204–2212.

[254] L. Carlone, S. Karaman, Attention and anticipation in fast visual-inertial navigation, IEEE Transactions on Robotics 35 (1) (2018) 1–20.

[255] A. Hussein, M. M. Gaber, E. Elyan, C. Jayne, Imitation learning: A survey of learning methods, ACM Comput. Surv. 50 (2) (Apr. 2017).

[256] B. D. Argall, S. Chernova, M. Veloso, B. Browning, A survey of robot learning from demonstration, Robotics and autonomous systems 57 (5) (2009) 469–483.

[257] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, J. Peters, An algorithmic perspective on imitation learning, arXiv (2018).

[258] F. Codevilla, M. Miiller, A. López, V. Koltun, A. Dosovitskiy, End-to-end driving via conditional imitation learning, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2018, pp. 1–9.

[259] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, Mastering the game of go with deep neural networks and tree search, Nature 529 (7587) (2016) 484–489.

[260] A. J. Ijspeert, J. Nakanishi, S. Schaal, Movement imitation

with nonlinear dynamical systems in humanoid robots, in: Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292), Vol. 2, 2002, pp. 1398–1403 vol.2.

[261] D. Silver, J. A. Bagnell, A. Stentz, Active learning from demonstration for robust autonomous navigation, in: 2012 IEEE International Conference on Robotics and Automation, IEEE, 2012, pp. 200–207.

[262] C. Dima, M. Hebert, Active learning for outdoor obstacle detection., in: Robotics: Science and Systems, 2005, pp. 9–16.

[263] N. Ab Aza, A. Shahmansoorian, M. Davoudi, From inverse optimal control to inverse reinforcement learning: A historical review, Annual Reviews in Control (2020).

[264] C. Daniel, O. Kroemer, M. Viering, J. Metz, J. Peters, Active reward learning with a novel acquisition function, Autonomous Robots 39 (3) (2015) 389–405.

[265] K. Judah, A. Fern, T. G. Dietterich, Active imitation learning via reduction to iid active learning, arXiv (2012).

[266] J. Ho, S. Ermon, Generative adversarial imitation learning, arXiv (2016).

[267] Y. Li, J. Song, S. Ermon, Infogail: Interpretable imitation learning from visual demonstrations, in: Advances in Neural Information Processing Systems, 2017, pp. 3812–3822.

[268] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, Advances in neural information processing systems 29 (2016) 2172–2180.

[269] K. H. Low, J. Chen, J. M. Dolan, S. Chien, D. R. Thompson, Decentralized active robotic exploration and mapping for probabilistic field classification in environmental sensing, in: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, 2012, pp. 105–112.

[270] I. Abraham, T. Murphey, Decentralized ergodic control: Distribution-driven sensing and exploration for multi-agent systems, IEEE Robotics and Automation Letters 3 (4) (2018) 2987–2994.

[271] G. Best, O. M. Cliff, T. Patten, R. R. Mettu, R. Fitch, Dec-mcts: Decentralized planning for multi-robot active perception, The International Journal of Robotics Research 38 (2-3) (2019) 316–337.

[272] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, J. S. Rellermeyer, A survey on distributed machine learning, ACM Computing Surveys 53 (2) (2020) 1–33.

[273] T. Li, A. K. Sahu, A. Talwalkar, V. Smith, Federated learning: Challenges, methods, and future directions, IEEE Signal Processing Magazine 37 (3) (2020) 50–60.

[274] A. D. Ames, K. Galloway, K. Sreenath, J. W. Grizzle, Rapidly exponentially stabilizing control Lyapunov functions and hybrid zero dynamics, IEEE Transactions on Automatic Control 59 (4) (2014) 876–891.

[275] A. D. Ames, M. Powell, Towards the unification of locomotion and manipulation through control Lyapunov functions and quadratic programs, in: Control of Cyber-Physical Systems, Springer, 2013, pp. 219–240.

[276] A. D. Ames, S. Coogan, M. Egerstedt, G. Notomista, K. Sreenath, P. Tabuada, Control barrier functions: Theory and applications, in: 2019 18th European Control Conference (ECC), 2019, pp. 3420–3431.

[277] A. D. Ames, X. Xu, J. W. Grizzle, P. Tabuada, Control barrier function based quadratic programs for safety critical systems, IEEE Transactions on Automatic Control 62 (8) (2016) 3861–3876.

[278] L. Wang, A. D. Ames, M. Egerstedt, Safety barrier certificates for collisions-free multirobot systems, IEEE Transactions on Robotics 33 (3) (2017) 661–674.

[279] F. Berkenkamp, M. Turchetta, A. Schoellig, A. Krause, Safe model-based reinforcement learning with stability guarantees, in: Advances in neural information processing systems, 2017, pp. 908–918.

[280] J. Choi, F. Castañeda, C. J. Tomlin, K. Sreenath, Reinforcement learning for safety-critical control under model uncertainty, using control Lyapunov functions and control barrier functions, arXiv (2020).

[281] A. K. Akametalu, J. F. Fisac, J. H. Gillula, S. Kaynama, M. N. Zeilinger, C. J. Tomlin, Reachability-based safe learning with Gaussian processes, in: 53rd IEEE Conference on Decision and Control, 2014, pp. 1424–1431.

[282] S. Bansal, M. Chen, S. Herbert, C. J. Tomlin, Hamilton-Jacobi reachability: A brief overview and recent advances, in: IEEE 56th Annual Conference on Decision and Control, 2017, pp. 2242–2253.

[283] I. Abraham, A. Prabhakar, T. D. Murphey, An ergodic measure for active learning from equilibrium, IEEE Transactions on Automation Science and Engineering (In Press).

[284] R. Cheng, G. Orosz, R. M. Murray, J. W. Burdick, End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 3387–3395.

[285] J. E. Marsden, T. S. Ratiu, Introduction to mechanics and symmetry: a basic exposition of classical mechanical systems, Vol. 17, Springer Science & Business Media, 2013.

[286] N. L. C. Chui, J. M. Maciejowski, Realization of stable models with subspace methods, Automatica 32 (11) (1996) 1587–1595.

[287] G. Mamakoukas, O. Xherija, T. D. Murphey, Learning memory-efficient stable linear dynamical systems for prediction and control, in: Conference on Neural Information Processing Systems (NeurIPS), 2020.

[288] B. Boots, G. J. Gordon, S. M. Siddiqi, A constraint generation approach to learning stable linear dynamical systems, in: Advances in neural information processing systems, 2008, pp. 1329–1336.

[289] W.-B. Huang, L.-L. Cao, F. Sun, D. Zhao, H. Liu, S. Yu, Learning stable linear dynamical systems with the weighted least square method., in: International Joint Conference on Artificial intelligence (IJCAI), 2016, pp. 1599–1605.

[290] N. B. Erichson, M. Muehlebach, M. W. Mahoney, Physics-informed autoencoders for Lyapunov-stable fluid flow prediction, arXiv (2019).

[291] N. M. Boffi, S. Tu, N. Matni, J.-J. E. Slotine, V. Sindhwani, Learning stability certificates from data, arXiv (2020).

[292] S. M. Richards, F. Berkenkamp, A. Krause, The lyapunov neural network: Adaptive stability certification for safe learning of dynamical systems, in: A. Billard, A. Dragan, J. Peters, J. Morimoto (Eds.), Proceedings of The 2nd Conference on Robot Learning, Vol. 87 of Proceedings of Machine Learning Research, PMLR, 2018, pp. 466–476.

[293] G. Mamakoukas, I. Abraham, T. D. Murphey, Learning stable models for prediction and control, IEEE Transactions on Robotics (Submitted).

[294] A. M. Mehta, J. DelPreto, K. W. Wong, S. Hamill, H. Kress-Gazit, D. Rus, Robot creation from functional specifications, in: Robotics Research, 2018, pp. 631–648.

[295] A. Berthoz, J. Petit, The physiology of action and phenomenology, Oxford University Press, 2008.

[296] J.-Y. Audibert, R. Munos, C. Szepesvári, Exploration–exploitation tradeoff using variance estimates in multi-armed bandits, Theoretical Computer Science 410 (19) (2009) 1876–1902.

[297] D. Ghosh, A. Gupta, S. Levine, Learning actionable representations with goal-conditioned policies, arXiv (2018).