Action representations in robotics: A taxonomy and systematic classification

Philipp Zech, Erwan Renaudo, Simon Haller, Xiang Zhang and Justus Piater

Abstract
Understanding and defining the meaning of "action" is substantial for robotics research. This becomes utterly evident when aiming at equipping autonomous robots with robust manipulation skills for action execution. Unfortunately, to this day we still lack both a clear understanding of the concept of an action and a set of established criteria that ultimately characterize an action. In this survey we thus first review existing ideas and theories on the notion and meaning of action. Subsequently we discuss the role of action in robotics and attempt to give a seminal definition of action in accordance with its use in robotics research. Given this definition we then introduce a taxonomy for categorizing action representations in robotics along various dimensions. Finally, we provide a systematic literature survey on action representations in robotics where we categorize relevant literature along our taxonomy. After discussing the current state of the art we conclude with an outlook towards promising research directions.

Keywords
Action Representations, Robotics

1 Introduction

In the beginning was the action[*] (von Goethe 1808, p. 81). Inspired by the Gospel of John, Goethe used this nowadays famous quotation in the third scene, first act of his famous play "Faust". Like Dr. Faust who back then struggled with a proper translation for the Greek word "logos", similarly we nowadays struggle with the exact meaning of the word "action". Despite various attempts at formalizing the notion of an action early in this decade, e.g., Davidson (2001) or Jeannerod (2006), the controversy on the exact nature of action is still active (see Section 2). Clearly, such a lack of understanding and of an accepted definition hampers research related to understanding human actions, e.g., in neuroscience or psychology, but also computational descriptions of action, e.g., in the field of robotics research.

Krüger et al. (2007) published a thorough review on action recognition and mapping in the fields of computer vision, robotics and artificial intelligence. They however stop short of providing a clear definition of action itself. Yet, Krüger *et al.* already provide a preliminary discussion of some criteria relevant for characterizing the notion of action. In our work, we build on these criteria (see Section 3).

More recently, Weinland et al. (2011) published a survey on vision-based methods for action representation, segmentation and recognition. Despite providing a thorough overview of existing approaches, their survey is limited to categorizing approaches according to their (i) spatial representation, (ii) temporal model, (iii) temporal segmentation, and (iv) view-independent representation. In contrast, in our work we aim to categorize action representations along many more dimensions (see Section 3). Further, Weinland *et al.* do not provide an underlying definition of action as a foundation for their classification. Last but not least, Weinland *et al.* do not consider the notion of an action's effect which not

only since Jeannerod (2006) is considered an integral aspect of an action representation but already dates back at least to Bernstein (1996).

The goal of our survey is to define classification criteria that are instrumental for a formal treatment of action representations in robotics. We thus aim at capturing the notion of action over a sufficiently broad range of analytical viewpoints that have emerged from both their theoretical interrogation but also from practical applications. We further present a thorough investigation of existing action-related research in robotics by categorizing relevant publications according to these criteria in a systematic way (see Section 4). As a result of this classification we then provide a comprehensive and qualitative discussion of existing research to identify both promising and potentially futile directions as well as open problems and research questions to be addressed in the future (see Section 6). To the best of our knowledge our work is seminal in both introducing a taxonomy for action representations in robotics and an in-depth discussion of existing research motivated by a quantitative study.

Contribution The core contribution of this article is the introduction of a comprehensive taxonomy for categorizing action representations in robotics. A systematic literature

Department of Computer Science, University of Innsbruck, Tyrol, Austria

Corresponding author:
Philipp Zech, Intelligent and Interactive Systems Group.
Department of Computer Science, University of Innsbruck
Technikerstrasse 21a
Innsbruck, Tyrol
Austria
Email: philipp.zech@uibk.ac.at

[*]Und schreibe getrost: im Anfang war die That!

search (see Section 4) of the keywords *action* and *representation* resulted in 1575 hits, which were systematically reduced to 469 considered papers. Out of those, we identified and categorized 152 major contributions in the field of robotics. For each publication it was possible to categorize the employed action representation as applicable. Given the resulting classification we then discuss the current state of the art of action representation in robotics (see Section 5). Finally, on the basis of this discussion, we identify promising directions for future research (see Section 6).

Intentional Limitations In this survey, only action representations that have an application in the field of robotics will be considered. Apart from that, we avoid categorizing papers that just build on existing models (see Section 4). Another limitation we impose on our survey is the deliberate exclusion of any papers or articles discussing plain controllers for implementing some movement. Though one could consider such a controller an action representation in some sense by arguing that it represents an "action" by its goal, i.e., a setpoint, we argue that controllers do not comprise an action representation simply by missing most of the aspects discussed in Section 3.

2   What is an action?

Despite being subtle in its form, the question of *what is an action* has a long history and probably first was investigated by Aristotle in his study on animal movement *De motu animalius*, where he contends that actions are justified as of a logical connection between goals and knowledge of effects (Russell and Norvig 2016; Nussbaum 1985),

> But how does it happen that thinking is sometimes accompanied by actions and sometimes not, sometimes by motion, and sometimes not? It looks as if almost the same thing happens in case of reasoning and making inference about unchanging objects. But in that case the end is a speculative proposition . . . whereas here the conclusion is which results from the two premises is an action . . . I need covering; a cloak is covering. I need a cloak. What I need, I have to make; I need a cloak. I have to make a cloak. And the conclusion, the "I have to make a cloak" is an action.

Aristotle pursued his studies further in his third book of the *Nicomachean Ethics* (Aristotle 1934). In his treatise—though now primarily focusing on ethics by attempting to answer the Socratic question of how men should best live—Aristotle already apprehended the imperative notion of human actions by attributing them a primary role in shaping a virtuous character. He thence introduces three categories of actions relevant to virtue, but also whether they are to be blamed, forgiven or even pitied:

- *Voluntary actions* are the righteous actions done by choice, i.e., on purpose. They result in increased happiness (*eudaimonia*).
- *Involuntary* or *unwilling actions* are neither praised nor blamed as in such cases no wrong action is chosen. This strongly builds on ignoring of what aims are good and bad.
- *Non-voluntary* or *non-willing actions* are bad actions done by choice, i.e., on purpose. They are preferred as all remaining options would be worse.

Admittedly, Aristotle did not discuss more specifically what an action is and also how it may be represented in our minds. Nevertheless, his thoughts are essential by clearly outlining different types of actions, thus ultimately implying that there must exist some internal representation which allows choosing among which action to do given a deliberate purpose. In contrast, if all actions are just hard-coded motor responses to external stimuli and no higher-level cognitive planning would precede action execution, such internal representations of actions would be pointless.

2.1   Action in psychology

In his article *Action-oriented representation*, Mandik (2005) discusses the nature of mental representations. Motivated by decade-lasting discussions between proponents of both underdetermined and determined (or active) perception, Mandik presents arguments from both conservative embodied cognition (CEC; or representationalism) and radically embodied cognition (REC) towards the nature of an internal representation of perception culminating in what he calls *action-oriented representation* (AOR).

Classically, the school of CEC calls for the need of an internal mental representation. This theory may be roughly identified as (Mandik 2005, p. 287)

> [. . .] the view that one has a perceptual experience of an *F* if and only if one mentally represents that an *F* is present and the current token mental representation of an F is causally triggered by the presence of an *F*.

Mandik then argues that the representationalist analysis of perception yields two crucial components: the *representational* component and the *causal* component. Whereas the former's job is to account for the similarity between perception on the one hand and imagery and illusion on the other hand, the latter is required to articulate the idea that in spite of similarities, there are crucial differences between perceptions and other representational mental phenomena (e.g., the relevant mental representation of an *F* must be caused by an *F* to count as percept of an *F*; Mandik 2005).

REC on the contrary argues against the explicit need for internal representations by relying on active perception. This essentially capitalizes on a perception-action cycle on the sensori-motor level in that actions are directly triggered by stimuli in the environment without the need for internal representations (*c.f.* Gibson 1966, 1979). Mandik argues however that active perception can be explained in terms of the representational theory of perception by acknowledging (Mandik 2005, p. 292)

> [. . .] that there are occasions in which outputs instead of inputs figure into the specification of the content of a representational state. I propose to model these output-oriented—that is, *action-oriented*—specifications along the lines utilized in the case of inputs. When focusing on input conditions, the schematic theory of

representational content is the following: A state of an organism represents *F*s if that state has the teleological function of being caused by *F*s. I propose to add an additional set of conditions in which a state can come to represent *F*s by allowing that a reversed direction of causation can suffice. A state of an organism represents *F*s if that state has the teleological function of causing *F*s.

Mandik then defines action-oriented representations (AOR) as any representation that has, in whole or in part, imperative content. Mandik thus argues that active perception—instead of rejecting the representational theory of perception—can contribute to the representational content of perception, and further, that percepts themselves may sometimes be action-oriented representations (Mandik 2005).

It is evident from Mandik's argument that internal mental representations are necessary for perceiving and understanding as well as interacting in the world. Further, it is obvious that these representations are required to subsume a certain amount of perceptual experience and action knowledge allowing an agent to plan for desired effects in the world. However, this still leaves us with our initial question of *what is an action*? What are the fundamental bits and pieces of both perceptual and sensori-motor experience that require internal symbolization to account for a mental representation of an action *A*?

Apart from Mandik, Jeannerod, in his famous book "Motor Cognition: What the Body Tells the Self" (Jeannerod 2006) provides an alternate treatment of action representations. First of all, Jeannerod argues that action representations must allow for mental simulation. Consequently, he distinguishes between *covert* and *overt* actions, where the former are the mental representations and the latter the actual, overt movements. He thus immediately attributes to action representations a functional nature (Vosgerau 2009), and hence argues that representing and executing an action is functionally equivalent. Secondly, Jeannerod states that actions are represented by their anticipated effect, that is, action representations essentially entail a mental model of a needed future environmental state. De Kleijn et al. (2014) further argue that such a representation in terms of an action's effects is unrenounceable as it unlocks contextualization of action control. This submission immediately relates to Jeannerod's third characteristic criterion of actions which is related to the actual type of an action. Jeannerod submits that there are two types of actions, viz. conceptual and non-conceptual actions. The crucial difference is that action representations with a conceptual content require an explicit representation of the goal, whereas for non-conceptual actions the goal is readily present in front of the agent and the action can be executed automatically without an explicit internal representation of the goal. This difference crystallizes in Jeannerod's example of intending to call someone via a phone. The first part of this action is to grasp the handset which clearly requires an internal representation of the goal—the phone itself—prior to executing the action. At the time of the execution however, the representation loses its explicit character and the remaining action, i.e., dialing, is executed automatically.

Similar to Mandik's treatise, it is also evident from Jeannerod's work that actions are internally represented. Contrarily to Mandik however, Jeannerod attributes to these representations a functional view by arguing that representing and executing an action is functionally equivalent. Whether one imagines or actually does an action employs the same neural substrates and processes (Jeannerod 2006). Jeannerod immediately provides a clear distinction between the resulting types of actions, i.e., conceptual and non-conceptual, as well as their manifestation, overt and covert, viz. being actually executed or just imagined.

2.2 Action in philosophy

Independently of the discussions in psychology, philosophy—most notably Donald Davidson with his philosophy of action—was looking for an answer to the question of *what is an action*. Contrarily to CEC and REC however, he aimed at identifying the relevant bits and pieces that physically constitute an action, independently of its mental representation. According to Davidson, an action, in some basic sense, is something an agent does that was *intentional under some description* (Davidson 2001). Davidson discusses this proposition in his famous example of someone accidentally alerting a burglar by illuminating a room, which she does by turning on a light, which she does by flipping the appropriate switch. Davidson is then concerned with the relation between the agent's act of turning on the light, her act of flipping the switch, etc., to answer the question which configuration of events, either prior to or contained within the extended causal process of turning on the light, really constitutes the agent's action. It is clear that there exists no unique answer to this question. Yet, the discussions caused by Davidson's example provide some insight into what may comprise an action. One may for example favor the overt arm movement that the agent performs, or the initiated causal process, but also the event of trying that precedes and "generates" the rest, i.e., the overt action. If for one second we stick to the latter definition of action, i.e., the mental act of trying, according to O'Shaughnessy (1997), this implies *willing*. Now according to O'Shaughnessy, an action then is defined as this mental act of willing which subsequently causes neural activity, muscle contractions and an overt actuation; happenings in the environment are just effects in the extended causal chain but not part of the action anymore. This however stands in stark contrast to De Kleijn et al. who submit that actions *are events that unfold in time and that must be structured in such a way that their outcome satisfies current needs and goals* (De Kleijn et al. 2014). Clearly, such a planned execution requires effects to chain the various deliberate events together.

2.3 Action in neuroscience

From a biological perspective, neuroscientists tried to link action with the neural substrates that generate it. These studies belong to the more general research on the production of task-adapted serial behavior in human beings. We summarize here the results from a roboticist's perspective

but for in-depth studies on action representation and neural substrates of motor control, see Grafton et al. (2009) and Hardwick et al. (2017) among others.

Researchers initially suggested that the hierarchy in information related to action (i.e. the goal constrains the motor programs to be executed) was reflected by a hierarchical organization of the brain areas. Keele and Jennings (1992) used serial reaction time tasks in combination with attention to assess sequence learning. Their results suggest that learning is easier when structure exists in the sequence, implying that the learnt representation relies on the combination of elementary patterns ordered given the task, hence some hierarchy.

Grasping studies also highlighted the influence of abstract information on motor execution. Jeannerod (1984, 1986) highlighted the interdependency between the formation of the grasp and the reaching movement, the latter depending on the former, whereas Rosenbaum et al. (2001, 1992) highlighted how the the hand shape of the grasp depends on the geometry of the object, how the tool will be used and how comfortable is the final posture.

Computational models have included action representation with both explicit (Cooper and Shallice 2006) and emergent hierarchy (Botvinick 2008) and successfully explained behavioral results. However, these models stayed at a representational level and did not directly adress the question of which neural substrates support the representation of action itself. A first proposition by Fuster (1999) tried to map anatomy with the expected hierarchy in the action representation. Imaging studies (Roland et al. 1980a,b) showed that motor cortex is only active during real movement execution whereas the supplementary motor area (SMA) is active during both executed and imagined movement. These results were interpreted as a sign that motor cortex and SMA play a role at different levels of abstraction and thus support the anatomical/functional hierarchy hypothesis.

However, several arguments come in opposition of a direct mapping between anatomy and functional hierarchy. We focus here on two of the four developed by Grafton et al. (2009, p. 643). First, a hierarchical model assumes a clear separation between the different levels and that only the lowest level is in charge of producing movement. However, it has been shown that even higher-level areas (premotor and parietal cortex, extrapyramidal brain stem pathways) project to the spinal cord and thus potentially influence the movement (Dum and Strick 1991, 1996). Secondly, the conceptual implication of a strict anatomical hierarchy raises the problem of the homonculus: if there is a decisional component on top of the architecture, this component itself may be organized hierarchically including a decisional component, etc. The resulting model would be complex which does not fit with the results on how fast and adaptable the action decision-making process actually is (Desmurget and Grafton 2000).

More recent studies of the anatomy have highlighted the existence of multiple parallel parietal-premotor-prefrontal loops in the brain. These loops seem to integrate multimodal sensory information rather than being tied to one modality only. They have been associated with object-centered action, tool use and reaching (Johnson and Grafton 2003; Rizzolatti and Luppino 2001; Rizzolatti and Matelli 2003). Grafton et al. suggest that the hierarchy of action representations is thus not tied to the anatomy itself but rather that (Grafton et al. 2009, p. 643)

> [...] an anatomical organization with multiple parallel parietal-prefrontal and premotor pathways supports a multitude of relative hierarchies that can be flexibly recruited as a function of task demands, experience, and context. In this framework, there are dissociable functional anatomic substrates, but these are not constrained by a fixed hierarchy. This shifts the focus of inquiry to understanding representational hierarchies that are highly flexible and goal based.

This second hypothesis has been investigated by focusing on the goal representation in motor execution studies involving grasping and bimanual coordination tasks. Grasping tasks directly map the goal to the target object, thus the task can be reframed as the problem of finding the proper transform between the perceived object and the hand. The anterior IntraParietal Sulcus (aIPS) in the parietal cortex has been shown to be critical for computing these sensorimotor transformations. The problem is then how the transformation information and goal representation are merged, that is, how does the aIPS perform the sensorimotor integration of the information?

Due to its connectivity to aIPS, the ventral premotor cortex is supposed to hold the goal representation. The hierarchical anatomy hypothesis would suggest that the sensorimotor information related to the target object is transformed into a goal representation. However, the hypothesis of a flexible hierarchy suggests that aIPS merges the sensorimotor and goal information and produces the constraints on the motor commands. This is supported by transcranial magnetic stimulation (TMS) studies (Tunik et al. 2005). Tunik et al. studied reaching and grasping tasks where the target object orientation (thus the goal) was changed very fast. The TMS was shown to disturb the ability of subjects to adapt to changes of the goal. The TMS blocks not only the adaptation of the grasp aperture but also the arm orientation. The authors claim that these results are better explained by the fact that aIPS does sensorimotor integration of the goal information rather than that TMS disrupts lower motor processes such as grip aperture. Consistent results are found in bimanual coordination: the change in the task goal changes the amplitude of the neural activity but does not change which regions are activated. Hence, there are areas (ventral premotor cortex and anterior intraparietal sulcus) in charge of maintaining the goal information, consistently recruited over tasks, that, when disturbed, have an effect on the adaptation of movement.

A similar dichotomy is shown in action observation tasks: Using the fMRI adaptation phenomenon (Repetition Suppression or RS), Hamilton and Grafton (2006) were able to show that the left aIPS is sensitive to which object is grasped (thus the "goal" of the action) whereas the information on the object position produces RS in other parts of the brain. They interpret this double dissociation as a result in favor of hierarchy between the goal of the action and the kinematic information of the action. In further

studies, they manipulated the shape of the grasp (Hamilton and Grafton 2007) or the outcome of actions (Hamilton and Grafton 2008) and were able to highlight segregated RS effects in specific areas of the brain. In the end, they argue that (Grafton et al. 2009, p. 648)

> [...] together, these three experiments support a model of representational hierarchy that distinguishes action means, kinematics, object-centered behavior, and ultimately, action consequences. The decoding of object-centered action appears to be strongly left lateralized, whereas the decoding of more complex action intentions arising as a consequence of the action engaged bilateral frontal-parietal circuits.

Actions are thus not uniquely represented in the brain but the representation is rather generated by the recruitment of several areas, with an apparent distinction between the goal-level information and the motor-related information. Moreover, Hardwick et al. (2017) recently did a meta-analysis on more than a thousand works from the literature on motor imagery (the mental rehearsal of an action), action observation (observing others' action execution) and movement execution (the overt interaction in the environment). They identified a consistent recruitment of a network of cortical or subcortical regions for each function. Both motor imagery and movement execution recruit the putamen which is involved in movement regulation. The body representation, encoded by the cerebellum, is also involved in motor imagery and movement execution along with the anterior and posterior midcingulate cortex for motor control. Action observation however does not recruit subcortical structures. It recruits the premotor parietal and occipital regions but less than during motor imagery.

These results from biology should teach roboticists two main lessons:

- The outcome of an action is a crucial part that defines it. There are dedicated areas to encode the goal and use the goal information to constrain the movement. Thus, an action in robotics should be defined by the goal it is intended to achieve, that is, its expected effects. The production of movement is then adapted to this goal. Thus, robot controllers should be flexible rather than reproduce stereotypical motions.
- Action requires multiple types of information that are not encoded in a central representation but rather distributed over and shared among multiple brain areas depending on the functional goal. For robotics, this argues in favor of a flexible representation of an action that links goal, movement and the currently-perceived scene.

Summarizing the above discussion clearly shows that despite being a core aspect of mammalian behavior, today we still lack a precise answer to the question of *what is an action*. Yet, this discussion however also shows that actions (i) are internally represented (*c.f.* Rizzolatti and Luppino 2001; Rizzolatti and Craighero 2004), (ii) are tightly bound to perception as a genuine source of information for action selection (Tunik et al. 2005), and (iii) yield effects which play a crucial role in shaping one's behavior (Hamilton and Grafton 2008).

2.4 Action in robotics

The notion of action occupies a paramount role in robotics. This simply stems from the circumstance that in order to meaningfully and intentionally interact with the world a robot requires knowledge about when to apply a specific action in order to achieve desired effects in the world. As Newton writes in her recent work on understanding and self-organization (Newton 2017, p. 5),

> Understanding is tightly coupled with the need of a living organism to take action. Understanding involves knowing how we might perform goal-directed actions relative to the environment. The experience of understanding is a feeling that the action affordances of a situation are not entirely unclear. Action (as opposed to reaction) requires imagery, including motor imagery, that can be used in the guidance of action.

Clearly, appropriate action representations are thus paramount for bootstrapping the development of an understanding of the world and ways an autonomous agent can meaningfully interact with this very world.

This paramount role of action representations was already pointed out by Krüger et al. (2007). In their survey they discuss the meaning of action at different levels in robotics from plain low-level sensory observations to high-level cognitive recognition and planning tasks. Krüger *et al.* argue that in order to nail down the meaning of action in robotics needs to address several areas, viz. observing and imitating others, control of one's own body, and learning of affordances (Zech et al. 2017). Their subsequent discussion provides an initial but yet unsatisfying answer to *what is an action*. However, we can clearly see that perception, embodiment, actuation and goal representation are core aspects of actions. We thus conjecture that such information requires a representation in order to be recallable. On the other hand it is necessary to talk about representations in the context of robotics as symbolic information, i.e., representations of knowledge, is crucial for computation. Aligned with the above discussion, we propose the following seminal definition of the notion of an action from a roboticist's stance in the next section.

2.5 A Seminal definition of action from a roboticist's stance

Motivated by the discussions so far we define the notion of an action for robotics as

- something an agent does that was intentional under some description,
- is caused by both the agent's current internal state and external percepts,
- is adaptive and deterministic to achieve desired effects,
- is learnt and symbolized while observing and imitating other agents,
- is mechanically effective,

- and primarily represented by its anticipated effects, that is, the goal.

Clearly, this definition is not final. However, we claim that it provides an initial basis for discussing what information, and especially in which form, eventually is required in order to elicit a general representation of actions for robots. It is obvious that perceptual aspects play a crucial role by virtue of the mutual relationship between perception and action (Bamert and Mast 2009). Further, learning plays an important role. Analogously to human development, one of the long-term goals in robotics research is to equip agents with robust learning capabilities about their environment and their own embodiment. Learning new means to interact with the environment, i.e., new actions, is paramount as not all situations an autonomous agent will experience are predictable. Thus, whereas providing initial knowledge about action bootstraps an agent's autonomy, the capability to adapt motions related to actions and subsequently learn new actions from experience is necessary to allow the agent to achieve novel effects that go beyond its current experience. As highlighted in Section 2.3, this can be achieved by integrating observations and experience from early sensory areas to higher-order cortical areas (*c.f.* Hasson et al. 2015).

Another important aspect of actions is their mechanical effectivity by causing overt changes in the environmental state; lacking a mechanically effective nature reduces an action to a mere gesture (Hobaiter 2017). Last but not least, actions—at least in the context of robotics—require external information that can be symbolized internally for goal-driven, behavioral planning. As already pointed out by Steels (2003), action representations are inevitable for planning. Given this seminal definition, in the next section we introduce our taxonomy for action representations in robotics.

3 Classification criteria for action representations

Given our discussions from Section *2* we can now introduce our taxonomy and its classification criteria for action representations in robotics. Clearly, a sound notion of action is paramount in that its representation for a robot is successful. Motivated by this we thus define an action representation in robotics as the union of an underlying *action* model and a *computational* model. Consequently, the action model deals with perceptual, structural, developmental and effect-related aspects, that is, the nature and embodiment of actions. In contrast, the computational model addresses low-level, implementational aspects of the mechanics of actions. Figure *1* gives an overview of our taxonomy and its classification criteria.

Before now discussing the criteria from Figure *1* in detail in Sections *3.1* and *3.2*, we want to remark that if a criterion is not specifically addressed in a given publication, it is assigned *not specified*.

3.1 Action model criteria

Action model criteria serve to asses the underlying "mental" action model of an action representation regarding its perceptual, structural, developmental, and effect-related aspects.

3.1.1 Perception Perceptual aspects study the means by which an autonomous agent employs different aspects of perceptual input for recognizing and memorizing actions in the environment. This dimension is standing in reason when considering Mandik's claim that perception and action are tightly coupled (Mandik 2005). An even stronger argument towards this tight linkage is given by Tucker and Ellis (1998) in arguing that *seen objects automatically potentiate components of the actions they afford*. Thus, one should consider visual inputs as one of the main drivers demarcating representations of actions.

3.1.1.1 Selective Attention Selective attention is becoming more and more popular in vision research, not least because of the impressive success of Deep Q-Learning (Sorokin et al. 2015). Naturally, selective attention is an important process for early action selection (Cisek and Kalaska 2010). Further, it allows noise and irrelevant information to be filtered out, focusing on what is important and relevant, thus raising awareness of one's own actions and ultimately culminating in conscious motor control (Webb et al. 2016). Thus, selective attention is either present or not (see rows 1 or 7–11, and 2–6 or 16–39, respectively, of Table **??**).

3.1.1.2 Granularity The granularity of the perceptual aspects of an action are important when it comes to generalizing actions. Clearly, in the context of a scene, actions can be perceived at different levels of granularity:

- *local* implies that an action model only considers local information, i.e., the part of an object that is relevant for doing the action like the handle of a hammer. As in the case of the perspective (*c.f.* Section 3.1.1.3) this comes with both advantages and disadvantages. For example, the agent may be capable of immediate interaction with the object upon recognizing a part but may fail to generalize its knowledge to different situations due to the lack of additional semantic information regarding the context in which the action is performed (see rows 34, 69 or 72 of Table **??**).
- *meso* implies that an agents perceives an action at the level of complete objects instead of only specific parts. This immediately allows an agent to acquire additional semantic information on the object itself enabling easier generalization of an action to different contexts as the agent has a more elaborate idea of what it can and cannot do with an object (see rows 1–2 or 35–38 of Table **??**).
- *global* implies that an agent perceives an action at the scene level. That is, not only does it perceive the concrete movements and objects involved but is also able to perceive the environmental context in which the action is performed, thus enabling consideration of interactions in the environment. Clearly, this allows an agent to easily generalize actions to novel contexts as it has acquired a complete picture of the circumstances under which an action can be performed. Observe however that this level of granularity does not readily imply generalization of the action (*c.f.* Section 3.1.2.4; see rows 3–4, 9 or 11–12 of Table **??**).
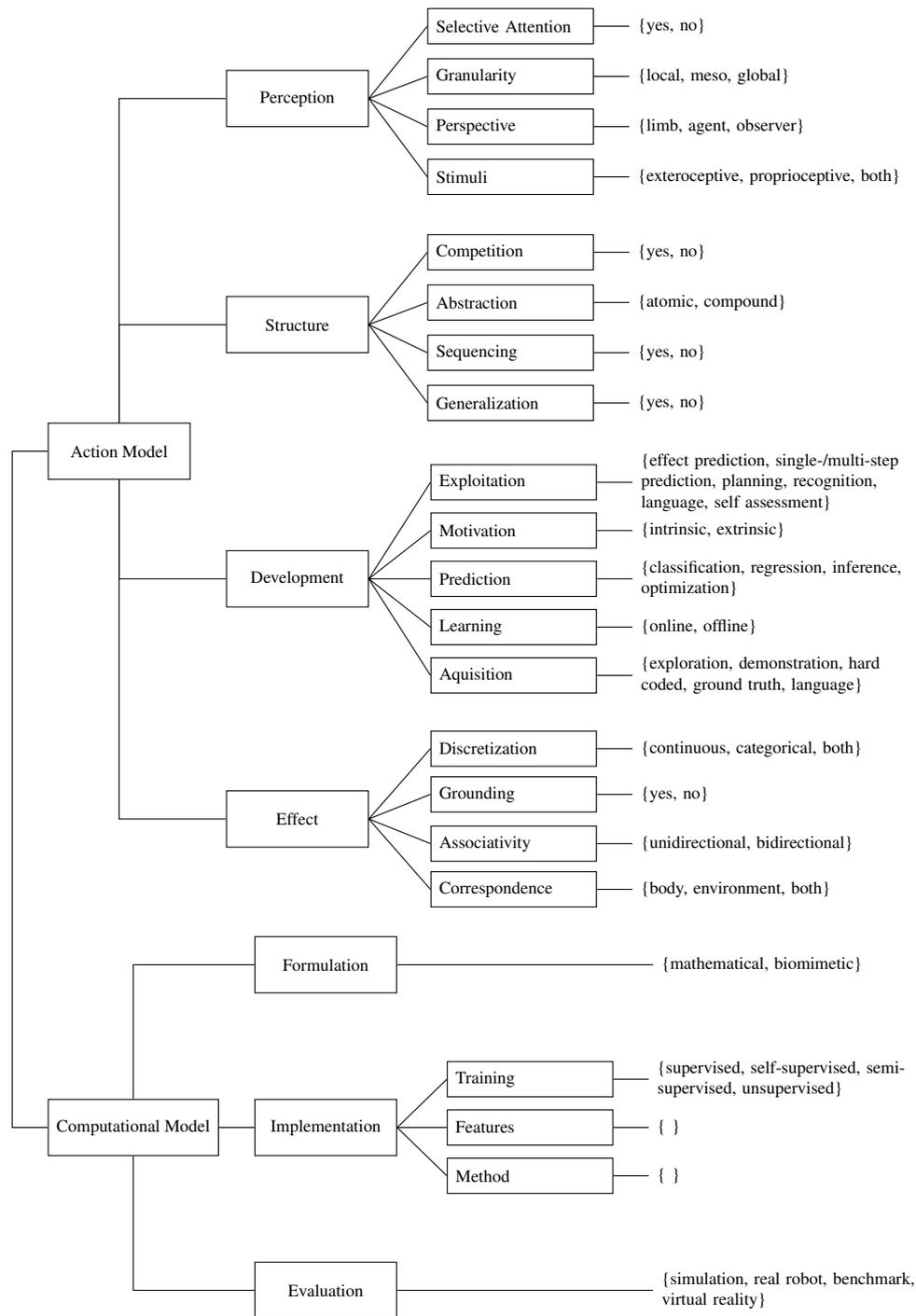
Figure 1. Overview of our taxonomy for categorizing action representations in robotics. For the sake of clarity, the choice *not specified* is excluded.

3.1.1.3 Perspective The perspective eventually nails down the reference frame of the perceived action. In the case of autonomous agents, multiple perspectives may apply given how the agent perceives and memorizes an action. We claim that there are three relevant perspectives autonomous agents can employ:

- *limb* implies that an agent learns actions with respect to one of its limbs, e.g., an arm or the end-effector only. The rationale is that our limbs are the primary means of interaction with the environment. This perspective

has the advantage that an agent may easily plan and adapt its actions locally, however may fail to do so at a global scale (see Section 3.1.1.2). Observe that this choice may imply the need for selective attention to properly isolate observations (see Section 3.1.1.1; see rows 27, 38 or 42 of Table **??**).

- *agent* implies that an agent perceives actions with reference to its whole body. This clearly has the advantage that an agent is able to plan and redo actions at a scale relevant for his body, yet it may

fail to capture fine-grained local aspects of an action. Compared to *limb* this choice usually refers to whole-body actions (see rows 2 or 5–6 of Table **??**).

- *observer* implies that an agent learns actions by observing them and associating them to the frame of reference of the agent executing the action, e.g., agents perceive actions from a third-person perspective. Clearly, the resulting action is represented at a global scale, yet the agent is required to—prior to execution—map the action into its own reference frame (see rows 22–24 or 30–32 of Table **??**).

3.1.1.4 Stimuli Stimuli, either external or internal, play an important role for action learning and representation as they encode relevant information that (i) triggers, (ii) monitors, (iii) allows adaption of an action both prior and during execution. Clearly, such stimuli may have different sources, e.g., internal or external. This criterion thus considers two types of stimuli:

- *proprioceptive* stimuli which relate to stimuli that are produced within the agent and its embodiment, e.g., force readings. Such stimuli are essential in that they enable monitoring the self during action execution (see rows 72–73 or 83 of Table **??**).
- *exteroceptive* stimuli which relate to stimuli that are generated in the external environment, i.e., interaction possibilities in the environment (affordances). Such stimuli are necessary for an agent to perceive the effects of its actions in the environment and subsequently replan or perform online adaptation of its movements to achieve its intended goals (see rows 30–38 or 40–60 of Table **??**).

Observe that this is a multi-choice criterion, i.e., an agent may as well consider both proprioceptive and exteroceptive stimuli for establishing an action model (see rows 2 or 5–7 of Table **??**).

3.1.2 Structure Structural aspects of the action model discuss the capacities of the representation in terms of cognitive capabilities it opens up to an agent. They are crucial for planning and reasoning for action selection in any given context. From an environmental perspective, structural aspects additionally discuss how the actions are organized in the environment.

3.1.2.1 Competition Obviously there may not always exist a single action that achieves an intended effect but instead a variety of actions equally allowing an agent to reach its goal, i.e., multiple actions are equivalent in terms of their effects but differ in their overt manifestation. To be able to select the ideal action, an action model is thus required to allow for competition among actions such that the agent may always choose the most suitable and efficient action. However, we do not attempt to study the internals of action competition but rather whether a model allows for it or not. Thus, Competition is either present or not. (see rows 1–7 or 10–14, and 8–9, 15–16 or 18–19, respectively, of Table **??**).

3.1.2.2 Abstraction Traditionally an action is considered atomic by triggering a specific movement applied in a specific context to achieve an intentional effect. However, considering actions only at such an atomic level subsequently hinders an agent to plan in terms of action sequences composed of a set of atomic actions. Our taxonomy thus considers both of these levels of abstraction as this ultimately enables an agent to reason in terms of higher-level actions and their goals:

- *atomic* actions encapsulate a single intentional effect. Atomic at this implies that an action cannot be further decomposed into smaller actions. Observe however that this does not restrict an atomic action to consist of a series of movements. For example, opening a drawer requires placing the gripper by moving the arm towards it, closing the hand around the handle, and subsequently retracting the arm (see rows 1–7 or 9–22 of Table **??**).
- *compound* actions on the contrary are actions that themselves consist of multiple atomic actions. That is, compound actions describe sequences of actions where these actions are combined and conditioned on their intermediary, intentional effects. Similarly to atomic actions, the agent usually aims at achieving again a single intended effect, yet at a larger timescale (see rows 23, 59–60 or 63 of Table **??**).

Observe that this is a multi-choice criterion, i.e., an agent may as well consider both atomic and compound actions when building its internal repertoire of action models (see rows 8, 58 or 100 of Table **??**).

3.1.2.3 Sequencing Being able to sequence actions eventually allows an agent to join both atomic and compound actions to reason about higher-level action goals and to achieve a variety of intended effects. Yet, we want to clarify that sequencing of actions does not readily imply that an agent is able to represent compound actions (see Section 3.1.2.2). Sequencing solely refers to the ability to generate long-term plans that may yield a variety of effects. Further, this criterion by no means studies the means of sequencing. Thus, sequencing is either present or not (see rows 1, 5 or 39–40, and 2–4 or 29–34, respectively, of Table **??**).

3.1.2.4 Generalization One of the most crucial aspects of autonomous robots is the capacity to generalize acquired knowledge to novel situations. Clearly, such a capacity places demands on the action representations. What would be the benefit of learning an action if it cannot be generalized to novel situations? Our taxonomy thus also studies this aspect of action representations as it holds a crucial factor for the success of an action representation. Again, however, we are not interested in the actual means of generalization at a computational level but just in whether the model allows it or not.Thus, generalization is either present or not (see rows 1–21 or 23–60, and 84, 104 or 132, respectively, of Table **??**).

3.1.3 Development Developmental aspects of an action relate to the means by which an agent is able to process new information to extend its action knowledge. Observe that this dimension is strongly tied to the perceptual aspects (see Section 3.1.1) of the action model in that the percepts ultimately constrain what can be learned. However, contrary to perceptual aspects which study *how the agent perceives* the environment for interacting with it, developmental

aspects study *how the agents learns to interact* with its environment.

3.1.3.1 Exploitation Available action knowledge can be exploited in different ways. However, different ways of exploiting one's knowledge result in different ways of how one subsequently interacts with the environment. Over the last decades roboticists have studied different ways of exploiting action knowledge where the range varies from selecting actions for reactive behavior to reasoning about actions for higher-level cognition:

- *effect prediction* of actions is an important capacity for autonomous agents as it allows them to understand both their environment but also their embodiment in terms of what they are capable of achieving. Additionally, effect prediction is a precursor for planning at large timescales (see rows 25, 64 or 76 of Table **??**).

- *single-/multi-step prediction* enables agents on the grounds of their immediate percepts and motivation to first search applicable actions and subsequently sequence them together given the predicted effects, or just to execute the most suitable action (see rows 1–2, 6 or 9–12 of Table **??**).

- *planning*, in contrast to single/multi-step prediction, cannot be done by exhaustive search. Rather, planning is implemented by reasoning over symbolic representations of both the environment and the agent's percepts and motivation, as well as its internally-symbolized action repertoire (see rows 13, 19 or 22 of Table **??**).

- *recognition* of actions and activities of others is crucial for autonomous agents that are supposed to help in our daily lives. Observe that this choice relates to effect prediction, yet at a different level. Whereas effect prediction ultimately allows an agent to predict what was the intention, action recognition allows an agent to already reason about how to achieve the intended goal instead of just capturing the sole intention (see rows 3–4, 7–8 or 16–18 of Table **??**).

- *language* enables agents to communicate with other agents by an important high-level cognitive ability. Agents exploiting their action knowledge by language ultimately are capable of communicating this knowledge in order to instruct others by means of teaching. Similarly, agents can also learn from spoken instructions (see Section 3.1.3.5; see row 52 of Table **??**).

- *self-assessment* of one's own capabilities unlocks to an autonomous agent the possibility of reasoning about its developmental state. This readily aligns with Jeannerod's famous idea that our actions tell us about ourselves (Jeannerod 2006). Further, being able to assess one's self and one's capacities and consequently knowledge gaps immediately allows one to tackle the exploitation vs. exploration trade-off by improving learned or acquiring new knowledge (*c.f.* Section 3.1.3.2).

3.1.3.2 Motivation Clearly an agent needs some kind of motivation that drives its process of knowledge acquisition. Such a motivation may either be external or internal. The former relates to external triggers, usually externally-imposed goals the robot is to achieve. The latter refers to internal motivations with no separable (clearly observable) outcome by an instrumental value (Ryan and Deci 2000). Consequently, this criterion has two possible choices:

- *extrinsic* motivation generally relates to external triggers that drive a robot to acquire new action knowledge. Observe that such extrinsic motivations may at some point overlap with intrinsic motivation (see below) in the case that an agent "realizes"—despite being externally imposed—that following some trigger may result in an overall improvement. In such an event we argue similarly to Ryan & Deci that this still should be considered external, as the original trigger is externally imposed (Ryan and Deci 2000; see rows 1, 26 or 33 of Table **??**).

- *intrinsic* motivation relates to internal triggers that drive the robot towards fostering or acquiring novel actions. The difficulty arising here is that robots generally are not able to deal with non-separable consequences like joy or satisfaction, which commonly are considered as triggers for intrinsically-motivated behavior (Ryan and Deci 2000). Yet, discussing this question is not the goal of our work, which is why we deliberately leave this question unanswered. Apart from that, intrinsic motivation has the disadvantage that the robot has to confront the exploration vs. exploitation trade-off, i.e., does it learn new actions or foster existing actions? On the contrary however, being intrinsically motivated enables an agent to learn what it is capable of and thus to develop an understanding of its embodiment (see rows 9, 106 or 152 of Table **??**).

Observe that this is a multi-choice criterion, i.e., an agent may be both extrinsically and intrinsically motivated in learning new actions.

3.1.3.3 Prediction After having learned new actions an agent needs the capacity to predict when a certain action is applicable (or required) given both its percepts and its motivation. Obviously, this criterion has a strong relation to the underlying computational model of our taxonomy (see Section 3.2) by relying on the mathematical tools employed. However, we argue that there still is a need for this criterion in the developmental dimension of our taxonomy, as properly deciding which action to take is a core aspect of developing sound and complete action knowledge:

- *classification* relates to agents which relate their perceptual input patterns to concrete categorical outputs. In this spirit, an agent identifies classes of actions which it implicitly relates to similar input patterns by defining a mapping from continuous to discrete spaces. Observe that classification transparently enables generalization (see Section 3.1.2.4; see rows 1–4 or 6–7 of Table **??**).

- *regression* relates to agents which estimate the proper action to take given relations in its perceptual inputs. That is, given its stimuli an agent learns a regression function that maps from continuous to continuous spaces (see rows 9, 11 or 15 of Table **??**).

- *inference* is a naturally inspired mechanism where an agent uses a set of acquired facts (existing knowledge) and hard-coded rules to infer new facts (novel knowledge), i.e., which action to take in a specific context. The rules may be represented as logical formulas, connections within graphs, or decision trees. Formally, this defines a mapping from discrete to discrete spaces (see rows 14, 19 or 31 of Table **??**).
- *optimization* is a purely mathematically-inspired mechanism to learn the best expected outcome given some input. Using it, an agent chooses an action that either maximizes a reward or minimizes a loss. Formally, this defines a mapping from either discrete or continuous to continuous spaces (see rows 5, 8 or 26 of Table **??**).

3.1.3.4   Learning Acquisition (see Section 3.1.3.5) of new information is an important capacity for autonomous agents to avoid stagnation. However, acquisition is only part of the deal. An agent also needs to be able to learn from this newly acquired knowledge in order to evolve. The means of learning are crucial for the development of both the agent and its internal action model. Our taxonomy studies this criterion by two possible choices:

- *offline* learning characterizes agents that first acquire data (or are provisioned with already-collected data) and subsequently employ this data for offline learning to acquire new knowledge. A drawback of this is that the agent may not be able to immediately react to changes in the environment or its embodiment, or to validate the learning outcomes itself in the real world (see Section 3.1.4.2). Yet, learning can be shaped more efficiently compared to online learning (see below; see rows 1–8 or 11–13 of Table **??**).
- *online* learning poses novel challenges to an agent, i.e., incomplete data and a large amount of noise and irrelevant data. That is, an agent, while exploring its environment to collect new data, is faced not only with the challenge to learn from this very data but also to filter out the relevant bits and pieces (*c.f.* Section 3.1.1.1). Despite this disadvantage, online learning comes with the advantage of immediate adaptability to changes in both the environment and the embodiment (see rows 9–10 or 14–15 or 25 of Table **??**).

3.1.3.5   Acquisition To be able to learn something new an agents needs to be provided with information it is able to process. Over the years, the robotics and machine learning community have drawn on various formats of information provision for agents. Clearly, each of those come with their unique advantages and disadvantages, which however are not the focus of this article. This criterion thus does not study advantages or disadvantages of the means of information provision but instead how the agent is provided with this novel information:

- *hard coded* implies that an agent generally does not acquire new knowledge but rather is provided with an initial set of, e.g., rules and facts about the world which allow it to shape its behavior. Clearly, such an agents stagnates until its knowledge base is manually extended (see rows 17, 48 or 52 of Table **??**).
- *ground truth* implies that an agent acquires new knowledge by learning to relate specific input stimuli to actual outputs (e.g., motor commands) for achieving a desired effect. Agents thus are able to learn but only if provided with valid feedback on their choices. Observe that ground truth traditionally is a manually-specified feedback signal that does not adapt to changes and may bias the learner (see rows 16 or 18–20 of Table **??**).
- *demonstration* implies that an agent learns from another agent or human teacher by being instructed on how to perform specific actions. This kind of acquisition comes with the advantage that the agent can immediately relate what it is shown to itself resulting in more efficient learning (see rows 1, 3 or 5–8 of Table **??**).
- *exploration* relates to agents that learn by exploring their environment by their own means, e.g., motor babbling. Being able to acquire new knowledge by exploring however requires the agent to be able to perceive and classify effects and changes in the world such that it can make sense of its actions (see rows 9, 14 or 33 of Table **??**).
- *language* probably is the most difficult but also most advanced means of acquiring novel action knowledge. The format may have lots of different variations, from direct imperative instructions (which are arguably the easiest to understand) to scene explanations from which the agent is required to extract the relevant bits and pieces that describe the action it is observing and is supposed to acquire. Clearly, being able to learn actions by language is an advanced, high-level cognitive ability and thus hard to achieve (see rows 22, 60 or 70 of Table **??**).

Observe that this criterion is again multi-choice, i.e., the means by which an agent acquires new knowledge are not restricted to just one source (e.g., an agent may learn about new actions by both being demonstrated what to do and at the same time being told what is actually done; see rows 11, 58 or 86 of Table **??**).

3.1.4   Effect As already claimed by Jeannerod (2006), in humans, actions are represented by their effects. Our taxonomy reflects this claim by containing a distinct dimension to study effect-related aspects of action models. Clearly, our notion of effect does not immediately correspond to a "mental" representation of an action. Nevertheless, it is an important aspect for studying the faithfulness of an action representation and its underlying action model.

3.1.4.1   Discretization Effect discretization studies the granularity of effect predictions that an action model supports. Effects may be either easily categorizable by clustering similar effects or they may reside in a continuous spectrum. In our taxonomy, the discretization of effects thence can fall into one of two categories:

- *categorical* effects generally relate to individual and different effects. Thence, effects under this

category generally describe fixed amounts or clearly-distinguishable events as a result of performing an action. Observe that both numeral and symbolic effects are subsumed by this choice (see rows 2–4 or 7–8 of Table **??**).

- *continuous* effects relate to fuzzy, boundless effects along a continuous dimension. Consequently, effects under this choice generally relate to real-valued action outcomes that are measurable along continuous spectra (see rows 5, 9–10 or 12 of Table **??**).

Observe that this is a multi-choice criterion, i.e., an agent may as well consider both categorical and continuous effects for establishing an action model (see rows 1, 5 or 37 of Table **??**).

3.1.4.2 Grounding Grounding of effects relates to the circumstance whether an action has or has not been executed in a real world environment by observing the intentional effects at the same time. Obviously, this criterion is of utter importance as it expresses the maturity of an action model. If once executed in a real-world setting with the intended effects observed, the action is both feasible and properly represented, whereas if not (i.e., only executed in simulation) one cannot guarantee that an action is actually doable as intended. Thence, grounding binds intended effects to observable real-world events. Thus, grounding is either present or not (see rows 1–2, 5–7 or 9, and 3–4, 8 or 10–13, respectively, of Table **??**).

3.1.4.3 Associativity Associativity of effects relates to the capacity of both predicting the effects of an action as well as predicting a necessary action to achieve a desired and intentional effect (Paulus et al. 2011). More precisely, this dimension does not directly investigate the mechanism for such capacities but instead whether the action model possesses this capacity and further, the nature of this capacity. Effect associativity can fall into one of two categories:

- *unidirectional* action-effect associativity categorizes an action model as only being able to infer the effects of executing a specific action. Consequently, an action representation lacks the capability of imagining which actions to execute to achieve a desired effect. On the contrary, given an action the model is readily capable of predicting the effects (see rows 2–13, 17 or 19–20 of Table **??**).
- *bidirectional* action-effect associativity categorizes an action model as possessing the capacity to predict relevant actions given some desired effect. This is ultimately related to mirror neurons which upon observation of an action (that involves and object) immediately activate neural populations relevant for motor control. This immediately allows for mental simulation of actions. However, observe that imagining does not readily trigger a representation (Elsner and Hommel 2001; Rizzolatti and Luppino 2001; Rizzolatti and Craighero 2004; see rows 1, 14–15 or 36 of Table **??**).

3.1.4.4 Effect Correspondence As argued by Newton (2017), usually we exercise an action to achieve a desired effect. Here we argue that one needs to distinguish between the actual frame of reference, or correspondence, of the effect. On the one hand, an effect may relate to changes in the environment, that is, displacing some object or opening a drawer. However, desired effects may also relate to changes in one's own bodily configuration, consequently treating the change in the environment as a consequence of the bodily change (*c.f.* O'Shaughnessy 1997; Section 2). Thence, the latter does not exclude changes in the environment but rather treats them as an indirect effect of executing an action triggered by the bodily effect. This criteria allows for three choices, viz. *environment* and *body* or the combination of both (see rows 4 or 9, and 1–2 or 5–7, and 3 or 8, respectively, of Table **??**).

3.2 Computational model criteria

Computational model criteria serve to assess implementational aspects of an action representation by how characteristics of the action model are realized. Thence, the computational model discusses the mathematical and theoretical underpinnings of action representations.

3.2.1 Formulation Here we consider whether a computational model is mathematically or bioglogically motivated. Clearly, there is a strong overlap between both categories, as, e.g., nature has inspired countless learning algorithms. Thus the question of where we draw the exact line between mathematical and biological motivation is valid. Our answer to this question is that a mathematically-formulated model solely draws on mathematical tools without the claim of being biologically plausible, whereas a biologically-inspired, or biomimetic, model aims at grounding its workings in biological and neural processes:

- *mathematical* implies that a computational model is purely relying on existing mathematical tools with no claim to be biologically inspired (see rows 1–11 or 13-20 of Table **??**).
- *biomimetic* implies that a computational model uses biology and cognition as a precursor for selecting proper mathematical tools. Such models thus are inspired from biology and neuroscience (see rows 12, 21 or 33 of Table **??**).

3.2.2 Implementation The implementational dimension of an underlying computational model of an action representation studies relevant aspects of the programmatic implementation. This subsumes (i) the concrete mathematical tools that are employed for learning and prediction, (ii) the environmental features that are used by the model, and (iii) the kind of training that is applied to the model, and thus entails a purely technical dimension.

3.2.2.1 Training The last dimension of the implementational aspects of the computational model of action representations studies the training used to train the predictive aspects of the developmental dimension of the action model (see 3.1.3). Our taxonomy supports the four most common types of training prevalent in robotics research:

- *unsupervised* learning relates to procedures where no – direct or indirect – feedback signal is used to drive the learning process. Eventually this requires an agent to detect relevant statistical patterns in

as well as the underlying structure of data without guidance. With respect to developmental robotics, this conceptually relates to the autonomous discovery of patterns or concepts from perceptual inputs in all available channels (exteroceptive and proprioceptive; see Section 3.1.1.4; see rows 3, 8 or 10 of Table **??**).

- *supervised* learning refers to learning given concrete feedback signals. That is, each input datum comes with a label informing the agent whether its prediction (or classification) was correct or not. Ultimately the agent learns to predict novel target values for previously-unseen inputs. Common drawbacks of this kind of training are under- or overfitting resulting from too little or biased training data (see rows 1–2 or 4–7 of Table **??**).

- *self-supervised* learning refers to agents capable of applying different views on data for learning patterns and concepts. Subsequently, one view, e.g., a specific sensor modality, is used to drive learning in another data view. For example, an agent may use clustering for learning low-level concepts in data (e.g., different obstacles). Subsequently, the cluster outputs are then used as target values for learning higher-level concepts using supervised learning (e.g., navigation). The term self-supervised refers to the supervision emerging from the learning agent instead of an external source (see rows 15, 27 or 73 of Table **??**).

- *semi-supervised* learning is a hybrid form of learning relying on techniques from supervised as well as unsupervised learning. It most naturally resembles human learning in that it is initially bootstrapped from supervised learning by a caregiver, followed by life-long, unsupervised learning by autonomous exploration (see rows 70, 87 or 111–112 of Table **??**).

3.2.2.2 Features To be able to make meaning of inputs in terms of computation, an action model requires extraction of features present in the inputs. Clearly, it may also directly rely on the inputs without any further processing. This criterion thus subsumes all kinds of representations from pixel intensities over salient points to features yielding from outputs of deep neural nets. Similar to the previous criterion this is also an open choice criterion, as again, the multitude of available and possible feature representations is too vast to be captured formally.

3.2.2.3 Method The method relates solely to the employed mathematical mechanisms that underpin the various perceptual, structural, developmental and effect-related aspects of the corresponding action model. It is an open choice criterion as providing choices for the multitude of mathematical tools that may be employed is too vast to be captured formally.

3.2.3 Evaluation The last dimension of the computational model underpinning an action representation discusses the means by which the action representation under study has been evaluated. The purpose of this dimension is two-fold: first, it indicates whether a model is just a theoretical musing or has practical relevance. Second, it indicates the maturity of a model. We thus claim that this dimension is of substantial importance. The choices are:

- *benchmark* refers to action representations that compete with others in terms of being evaluated on an unbiased, explicitly-devised data set. Doing so immediately allows comparing representations with each other in terms of their representational and functional capacity. Benchmarks can fall into two categories distinguished by how the baseline is established. In one case, the baseline is computed from a specially-devised training data set and compared against a test data set. In the other case, a baseline is established from the results of reference studies investigating the same hypothesis to be then compared against the own model using the same data as the reference studies (see rows 3–4 or 7–8 of Table **??**).

- *real robot* implies that an action representation has been evaluated on a real, physical robot. Clearly, this kind of evaluation is the strongest one as it requires a model to be robust against real-world noise and to be able to deal with potentially-incomplete data (see rows 1–2, 5–6 or 9–15 of Table **??**).

- *simulation* categorizes models as having only been evaluated in a simulated environment. Clearly, such an evaluation is weaker as the inevitable physics approximations and imperfect noise models fail to catch a real-world environment. Thus, for action representations only evaluated in simulation one cannot assess much more than that they may be practically feasible but not whether they truly are or not (see rows 21–22, 26 or 39 of Table **??**).

- *virtual reality* is a relatively recent type of evaluating, among others, action representations (Zech et al. 2017). It refers to a type of evaluation where a human agent provides non-simulated interactions in an otherwise simulated environment with a simulated agent (see rows 31 or 95 of Table **??**).

Observe that this is a multi-choice criterion, i.e., a computational model of an action representation may well be evaluated in multiple settings, e.g., preliminary evaluation in simulation with subsequent evaluation on a benchmark (see rows 91 or 110 of Table **??**).

4 Selection and classification of papers

Paramount to performing a systematic literature review together for categorizing papers is a carefully designed search and selection procedure. This section will thus introduce our systematic search and selection procedure for identifying papers relevant for classification. Additionally, we identify relevant threats of validity to our study. The resulting classification of action representations in robotics covered in the selected publications is then used in the next section to indicate the adequacy of the defined criteria and for further discussions (see Sections 5 and 6).

4.1 Selection of publications

The selection of relevant, peer-reviewed, primary publications requires the definition of a search strategy as well as paper selection criteria together with a selection procedure applied to the collected papers.

4.1.1 Search strategy The initial search conducted to collect candidate papers was done automatically on December 1st, 2017 by consulting the following digital libraries:

- IEEE Digital Library (`http://ieeexplore.ieee.org/`),
- ScienceDirect (`http://www.sciencedirect.com/`),
- SpringerLink (`http://link.springer.com/`),
- SAGE (`http://journals.sagepub.com`), and
- Frontiers in Neurorobotics (`https://www.frontiersin.org/journals/neurorobotics`).

These libraries were chosen as they cover most of the relevant research on robotics. The search string was kept simple, i.e.,

```
action representation AND robot
```

in order to keep the search general enough and to avoid missing any publications employing more precise terminology. Observe that the search was applied to all of the following search fields: (i) paper title, (ii) abstract, (iii) body, and (iv) keywords. The search produced a set of 1575 retrieved papers, thus a paper selection process was subsequently employed to further filter the results.

4.1.2 Paper selection Figure 2 summarizes the paper selection process which comprised three phases. In the first phase, papers were excluded based on their title: if the title did not indicate any relevance to robotics and action representations, papers were discarded from the classification. This reduced the initial set of 1575 papers to 686 remaining papers. In the second phase, papers were excluded based on their abstract, reducing the number of relevant papers to 469. In the third and final phase, papers were rejected based on their content, reducing the set of relevant papers to 152. Thus, our classification, as discussed in Section 6, includes a total of 152 papers. Note that during the last iteration, a number of relevant papers were rejected on the basis that they either failed to introduce a novel representation or to sufficiently reevaluate an existing representation. Further, we deliberately excluded papers focusing solely on gesture recognition, as these generally are not considered mechanically-effective motions compared to actions (Hobaiter 2017).

4.2 Paper Classification

The 152 selected publications were categorized according to the classification criteria as defined and discussed in Section 3 by four researchers. For this purpose, the remaining set of primary publications was randomly split into four sets of equal size for data extraction and classification. A classification spreadsheet was created for this purpose. Besides bibliographic information (title, authors, year, publisher) this sheet contains *classification fields* for each of the defined criteria. To avoid misclassification, the scale and characteristics of each classification criterion were additionally implemented as a selection list for each criterion. As explained above, the list also contained the

item 'not specified', to cater for situations where a specific criterion is not defined or could not be inferred from the contents of a paper. Problems encountered during the classification process were remarked upon in an additional comment field. The resulting classification of all publications was then reviewed independently by all four researchers. Finally, in multiple group sessions, all comments were discussed and resolved among all four researchers.

4.3 Threats to validity

Naturally there exist various issues that may influence the results of our study, e.g., the defined search string as discussed previously. Threats to validity include multiple factors, most relevant to us (i) publication bias, (ii) identification and (iii) classification of publications, as well as the (iv) terminology employed.

4.3.1 Publication bias This threat relates to the circumstance that only certain approaches, that is, those producing promising results or promoted by influential organizations are published (Kitchenham 2004). We regard this threat as moderate since the sources of publications were not restricted to a certain publisher, journal or conference. Therefore, we claim that our study sufficiently covers existing work in the field of action representations and robotics. However, to balance the trade-off between reviewing as much literature as possible while nevertheless accumulating reliable and relevant information, gray literature (technical reports, work in progress, unpublished or not peer-reviewed publications) was excluded (Kitchenham 2004). Further, the required number of pages was set to four to guarantee that publications contained enough information in order to categorize them appropriately.

4.3.2 Threats to the identification of publications This threat is related to the circumstance that, during the search and selection of publications, relevant papers may have been missed. To address this, we employed a very general search string to avoid missing potentially relevant publications during the automated search. Yet, to additionally reduce the threat of missing important publications, we informally checked papers referenced by the selected papers. We did not become aware of any frequently cited papers that were missed.

Apart from that, we also want to point out that we deliberately excluded any papers discussing just plain reactive open- or closed-loop controllers, e.g., Dynamic movement Primitives (DMP) or Central Pattern Generators (CPG), as these, to the best of our knowledge, do not readily address the topic of action at a cognitive level but rather at the control level. Clearly, reactive control does not relate to the cognitive concept of an action being represented in terms of its effects and usually not readily coupled to some specific motor program. Additionally, we also excluded a large number of papers studying the application of reinforcement learning (RL). In general, RL assumes actions are already given (observe that we are interested in action representations and means of populating them by learning), and further, reinforcement learning also does not employ any notion of effect whatsoever.
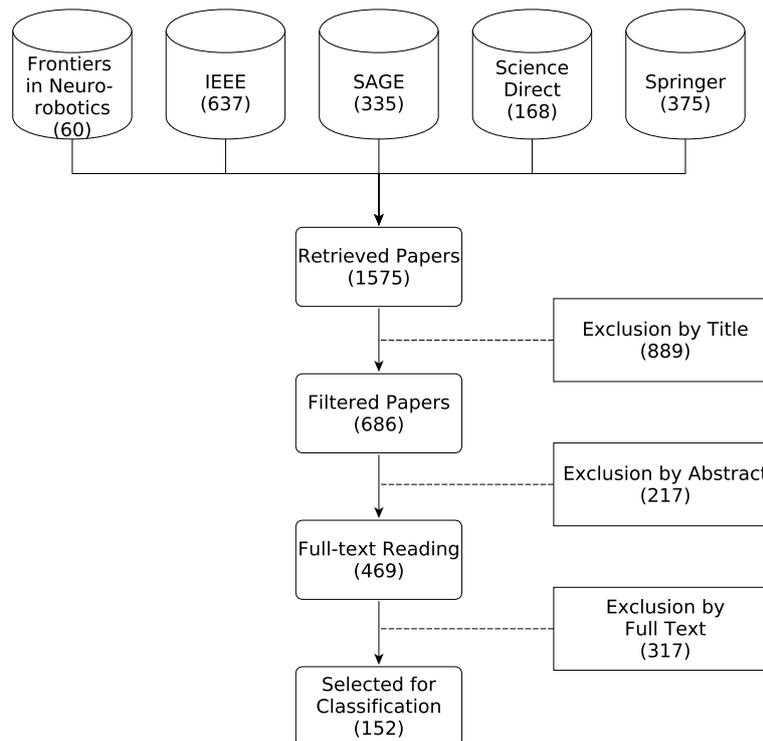
Figure 2. Selection of publications studied in this survey.

4.3.3 Threats to the classification of publications Given the rather large number of publications selected for classification according to a substantial number of defined criteria, the threat of misclassification needed to be addressed. Various measures were implemented in order to mitigate this threat. First of all, all criteria were precisely defined, as presented and discussed in Section 3, prior to the commencement of the paper selection and classification process. There was scope for the refinement of the concepts by the researchers during the process, but this was restricted mainly to descriptive adjustments. Secondly, for each of the criteria we added a list of possible selections in the classification sheet to avoid misclassification. Third, the classification was conducted in parallel by four researchers who are experts in the field and who repeatedly cross-checked the classification independently. Finally, weekly meetings were held by the four researchers to discuss and resolve any comments that arose during independent classification.

4.3.4 Terminology We are aware that the way we use specific terminology, e.g., *action* and *motion*, or, *learning* and *inference*, or *understanding* may not be perfectly in line with their use in other areas of research. However, this survey has been written with a robotics research background, which is why we stick to the terminology as used in this field. Thus, given both this circumstance and the fact that the notion of an action representation, at least for now, is not that wide-spread in robotics we took the liberty to rigorously decide on our own when to use which term and whether some representation is an action representation or not. However, readers from different fields should not face any problems in properly interpreting the content of this work, as the terminology as used in robotics research—to

a high degree—has been coined by relevant concepts from psychology and neuroscience. On the other side, we hope that our work stimulates a discussion about the state of the art of action representations in robotics to advance this field and contributes to the establishment of a common and well-defined terminology.

5 Results and Discussion

This section comprises the main contribution of this article by presenting and discussing the classification of the selected papers (see Section 4). The complete classification of all 152 papers by the introduced taxonomy (see Section 3) is shown in the appendix of this article (see Tables **??** and **??** in Appendix C) and is also available online[†].

For each of the selected publications it was possible to categorize the presented action representation according to the criteria defined in Section 3. This indicates the pertinence of these criteria for the classification of action representations in robotics, thence supplying a framework for understanding, categorizing, assessing, and comparing action representations in robotics. Additionally, besides validating the criteria introduced in Section 3, our classification, having been conducted in a systematic and comprehensive manner, provides an aggregated view and investigation of current state of the art of action representations in robotics.

Figure 4 shows the summary statistics by a co-occurrence matrix of category values as defined in Section 3 that arise in the analyzed papers, thus providing the foundation

---

[†]https://iis.uibk.ac.at/public/survey/
ActionRepresentation/

for subsequent discussions. Figure 3 gives the category distributions of the selected papers.

5.1 Learning of Action Representations

Learning, that is, the process of acquiring new or modifying existing knowledge, behaviors, skills, values, or preferences (Gross 2015), is one of the central aspects of action representations. Clearly, this usually requires proper motivation for learning to take place. Looking at Figure 3 shows that in great measure, the question of how to motivate (extrinsically or intrinsically) a learner is hardly addressed (30 out of 152) and where it is, learning is chiefly extrinsically motivated (26 out of 30). Correlating this with the kind of training (see Figure 4), we conjecture that this in general is because of the prevalence of supervised and offline learning (79 and 110 out of 152, respectively) which traditionally imposes the motivation of reducing some externally prescribed loss. In accordance to that, exploratory learning has also seen very little attention (16 of out 152). Clearly, such kind of learning would require switching to semi- or self-supervised online learning (1 and 4 of out 16 that do online learning). Furthermore, doing so would require a valid model of a robot's embodiment in order to learn what is possible given the available motor skills. In line with this, we also argue that manually-provided ground truth should be avoided as a means of a feedback signal for learning due to its static nature (69 out of 152). Using such manually-defined ground truth drastically impedes autonomous learning on a real robotic platform due to the dependence on teacher-dependent supervision (54 out of 69). Again, if learning is done on a real robotic platform we suggest the use of semi- or self-supervised online learning for immediate relation to the robot's embodiment.

The majority of the considered methods uses the observer perspective (86 out of 152). Clearly, learning from such a perspective hinders the emergence of action representations for purposes other than plain recognition due to the yet-unsolved correspondence problem (*c.f.* Zech et al. (2017)) and the consequent difficulty of relating observed actions to one's own embodiment. Admittedly, one can learn from observation but only in combination with subsequent exploration. Yet, we did not identify any such paper. On the bright side, however, there is still a substantial number of approaches that learn from the agent's perspective (55 out of 152), though only 14 of those acquire new knowledge by exploration and 20 by demonstration. This readily corresponds to the prevalent use of only exteroceptive stimuli (113 out of 152). Observe that this again drastically foils relation to an agent's own embodiment.

Noteworthy further drawbacks we currently see in learning action representations are (i) a lack of employing selective attention (27 of out 152), (ii) scarcity of language use (3 out of 152), (iii) negligence of learning with reference to an agent's limbs (10 out of 152), and (iv) only considering discrete instead of continuous (or both discrete and continuous) effects (78 out of 152). Obviously, selective attention allows the curse of dimensionality to be tackled by focusing on what is relevant. Further, learning with respect to the limbs eases re-execution of trained actions due to the simplified planning problem, i.e., there is no need to do whole-body planning. Thirdly, using language enhances

structuring and understanding of action knowledge thanks to the tight relation between language and action (Guerra-Filho and Aloimonos 2007). Finally, enabling agents to reason about not only discrete but also continuous effects unlocks the ability to plan with respect to local changes in both the environment and the embodiment, and not only at a global environmental scale.

To conclude, in the area of learning action representations, the current multiple drawbacks stem in general from the prevalent combination of supervised, offline learning from an observer's perspective. We suggest that in the future, online learning in a semi- or self-supervised way from the agent's perspective merits more emphasis to resolve issues like the correspondence problem or proper motivation for learning.

5.2 Maturity of Action Representations

Two of the central criteria of our taxonomy directly relating to the maturity of an action representation are the means of exploitation and evaluation. Clearly, representations that allow only for recognition and that further are only evaluated on a benchmark lack maturity, missing empiricism yielding from real-world experiments on an actual robotic platform. In this respect, Figure 3 draws a rather disappointing picture in that more than half of the categorized papers have only been evaluated in terms of benchmarks (80 out of 152). Correlating this to the type of exploitation (see Figure 4) we see that the bulk of these papers (72 out of 80) only do recognition. The main drawback coming along with such methods is the use of only exteroceptive stimuli and features which undermine construction of internal representations of one's own embodiment due to the missing relation between observation and embodiment. Yet, as neuroscience claims, such representations of one's embodiment are paramount for action recognition. Only with such models of the self at our disposal we are able to map observed actions onto our own embodiment for reexecution (Sokolov et al. 2010). This mapping is crucial as it immediately solves the the correspondence problem. On the other side, the nastiness of the correspondence problem in combination with lacking representations of the self (and thereof emerging relations to an agent's embodiment) immediately explains why the works that address action recognition fail to close the gap towards re-execution of observed actions.

Another problem that emerges if looking closer at the plethora of papers doing action recognition is their stopping short of action sequencing (0 out of 72 address action sequencing). However, looking at Figure 3 immediately reveals that papers not focusing on recognition but rather on single- and multi-step prediction as well as planning are capable of sequencing actions (27 and 16 out of 63). Unfortunately however, these methods only allow sequencing single actions together but fail to represent resulting action sequences as compound actions. In contrast, those action representations that are able to handle compound actions (4 out of 152) do not address sequencing of such compound actions.

The ability to handle action competition in the representation is another key aspect regarding the maturity of a model. Clearly, in every situation an agent is faced with multiple actions that yield similar or identical effects; thus it has to choose which action, among the feasible ones, to ultimately

Development

Motivation		Exploitation	
extrinsic	26	planning	20
intrinsic	4	single/multi-step prediction	46
not specified	122	effect prediction	6
Acquisition		language	1
demonstration	55	recognition	78
exploration	16	*not specified*	1
ground truth	64	**Learning**	
hard coded	9	online	25
combination	5	offline	125
not specified	3	*not specified*	2
Prediction			
optimization	21		
inference	18		
classification	78		
regression	33		
not specified	2		

Perception

Perspective		Abstraction	
limb	10	atomic	136
agent	55	compound	10
observer	86	both	6
not specified	1	**Competition**	
Stimuli		yes	61
exteroceptive	113	no	91
proprioceptive	11	**Sequencing**	
both	27	yes	63
not specified	1	no	88
Selective attention		*not specified*	1
yes	27	**Generalization**	
no	125	yes	146
Granularity		no	5
global	88	*not specified*	1
meso	55		
local	7		
not specified	2		

Effect

Discretization	
categorical	78
continous	48
both	9
not specified	17
Grounding	
yes	46
no	91
not specified	15
Associativity	
unidirectional	98
bidirectional	17
Correspondance	
environment	42
body	65
both	25
not specified	20

Formulation

mathematical	131
biomimetic	21
Training	
unsupervised	41
self-supervised	6
semi-supervised	8
supervised	94
not specified	3
Evaluation	
VR or Combination	4
Simulation	21
Real robot	46
Benchmark	80
not specified	1

Figure 3. Numbers of papers falling into each category for all criteria.

execute. In total, however, the number of approaches able to handle competition is less than half of all the papers we categorized (61 out of 152). Yet, using proper mathematical mechanics one actually can get competition for free, e.g., by employing neural networks or any other type of regressor/classifier that intrinsically handles competition at the decision level. However, we see a further potential reason for this general lack of handling competition, motivated by the circumstance that most works are only able to handle a couple of actions, possibly rendering competition useless for now. Yet, future work should put more emphasis on action competition, rendering agents more autonomous.

A last but very important indicator for the maturity of an action representation is the way it represents and handles effects. In general, the works we categorized focus on categorical effects (75 out of 152 papers) with only unidirectional associativity (95 out of 152). Correlating this to the category of exploitation, we again see that the majority of the representations that are only able to handle categorical effects are exploited for action recognition (54 out of 75). Clearly this is due to only recognizing classes of actions but not the continuous changes that the effects yield in both the agent's embodiment and its environment. However, this substantial lack of handling continuous effects has a further reason: a shortcoming in grounding effects in the real world (46 out of 152). Real-world physics in general are not discrete but continuous dynamical systems. Only by verifying estimations by real-world observations can we expect an agent to truly learn about the effects that it causes as well as its potential control over its environment. Finally, a last major drawback from our perspective is the prevalent unidirectional effect association (98 out of 152). This immediately yields scarcity of inverse models for inferring what to do to achieve a desired effect, consequently reducing the autonomy of the agent.

To sum up, we submit that the majority of existing action representations are not in a very mature state. This follows from three major observations. First, evaluation mostly is not done on real robotic platforms. Secondly, researchers for now mainly focused on constructing representations only for recognition that neglect the self. And third, for most of the categorized works effects are not grounded in real-world physical environments. By putting more emphasis on these issues we claim that existing drawbacks, e.g., the shortcoming of proper inverse models, could readily be addressed.

5.3 Formalizing Action Representations

One of the central yet quite disappointing insights of our systematic search and classification is the realization that in robotics, usually, there is no widespread use of specifically devised data types (think about an abstract data type) for storing and managing action-specific knowledge. Clearly, such data types are however necessary as our earlier treatise in Section 2 shows where in general one can see strong arguments in favor of internal representations of both actions and the self (*c.f.* Mandik (2005); Jeannerod (2006); Tunik et al. (2005); Naito et al. (2016)). Yet, except for the work of Beetz *et al.* (Tenorth and Beetz 2012; Bartels et al. 2013), as well as Wörgötter *et al.* (Worgotter et al. 2013; Aksoy et al. 2013; Vuga et al. 2015; Aksoy et al. 2016b) there has been little effort towards the design of appropriate data structures for storing, accessing, and transferring action knowledge. Quite the contrary, what is done in most categorized papers is to leverage existing vision-based feature extractors (e.g., CNNs) and descriptors, and to subsequently use a combination of those as input to some regressor/classifier. Obviously these vision-based features and descriptors in general do not express anything related to a specific action except for maybe what it "looks like", but doubtlessly no information regarding how to actually perform the action (*c.f.* our earlier writing on closing the gap between recognition and re-execution in Section 5.2). Apart from that, in respect of Searle's famous definition of a computer being *a device that manipulates formal symbols* (Searle et al. 1997), we conjecture that for artificial agents, valid representations of both actions and the self are inevitable. Formal symbols are representations. So, at the end of the day, an artificial agent needs internal representations to be able to compute.

Since Francis' influential article on the internal principle of control theory (Francis and Wonham 1976) it is generally accepted that one of the central pillars of mammalian motor

Figure 4. Co-occurrence matrix of all criteria for all categorized papers (best viewed on a computer display; numbers missing for each criteria to sum to 152: not specified).

cognition strongly builds on inverse models for motor control (Wolpert and Kawato 1998). In the course of our survey we identified exactly one paper out of 152 (see Table **??** and **??** in the Appendix) that makes use of explicit inverse models for single-/multi step prediction. Obviously this astonishing ignorance of inverse models only fortifies what we already argued earlier regarding the maturity of action representations. Yet, this lack of inverse models readily can be tackled by carefully revising existing representations and their mathematical underpinnings. We claim that doing so is paramount to verily advance the current state of the art in action representations in robotics. From a present-day perspective, in the long run this would also aid in effect modelling for action representations, as one readily obtains bidirectional effect associativity which currently is only addressed by a fraction of all categorized papers (17 out of 152). We guess that the concurrent absence of inverse models as just discussed is further fostered by also not attributing neuroscientific results enough consideration in terms of building biomimetic models for action representations (21 out of of 152).

Another blind spot we revealed in the context of formalizing action representations is that, to a great extent, model formalizations are only done at the subsymbolic level. That is, looking at Tables **??** and **??** one sees a strong predominance of methods that purely operate at a subsymbolic level by means of the used features. Clearly, higher-level cognition requires symbolization of acquired knowledge for high-level abstract task planning. The results of our classification as shown in Figure 3 reinforce our observation in that only a small fraction of categorized action representations are exploited for high-level task planning (20 out of 152). We argue that action representations require proper symbolization for unlocking high-level abstract task planning.

Finally, a last point to discuss in the context of action representation formalizations is the scant use of optimization (21 out of 152). We argue that optimization should be a first-class choice as ultimately one wants to optimize behavior by choosing the most fitting action. Correlating these papers to the kind of exploitation we at least see that 8 out of those do single-/multi-step prediction, and 9 do planning, respectively, indicating that if optimization is used, then it is for optimizing behavior. Nevertheless, we argue that more emphasis should be put on optimization for action selection and behavior shaping. Observe that this however does not call for an increased use of RL at this point. RL in general is not about optimizing an action but rather the sequence of actions that is taken to fulfill a task. Optimization of the action itself should take place before policy optimization.

5.4 Usability of Action Representations

One of the paramount questions when talking about formal models in a general sense is their usability. The Oxford English dictionary defines usability as *the degree to which something is able or fit to be used*. Now, this definition is very broad and does not really investigate what it means to be usable or how to actually measure whether something is usable. Let us therefore expand this definition by introducing three characteristics that we consider relevant for quantifying the usability of an action representation:

- *effectiveness*, i.e., the completeness and accuracy of a representation
- *efficiency*, i.e., how long does a representation need to be learned and also how easily can it be leveraged for executing a desired action
- *robustness*, i.e., how well does the representation generalize, but also deal with incomplete/corrupt data

Regarding effectiveness we clearly see a large shortcoming in currently-available action representations. Looking at Figure 3 (and as already mentioned) the bulk of existing methods solely do action recognition (78 out of 152). Despite being aware that recognition capabilities are crucial for action representations, we however claim that this is only the first step towards more powerful representations that also allow for motor imagery and actual execution of the abstracted action. Especially single- and multistep prediction is of high importance (46 out of 152) due to its immediate relation to deciding what to do next. Unfortunately however, this again boils down to closing the gap between recognition and execution (as already mentioned) as well as the correspondence problem for properly learning from demonstration. Further, this also comprises consideration of continuous effects for being able to come up with precise and accurate predictions regarding dynamic changes in the environment.

Regarding efficiency we submit that current models are learnable with reasonable expense, at least in the event of supervised, offline learning (85 out of 152). However, one has to keep in mind the general shortcoming of such models in that they generally only allow for action recognition (53 out of 85). Clearly, one has to keep in mind that in the case of exploratory, self- or semi-supervised learning, learning a representation will take substantially longer. Unfortunately, as our survey shows, exploratory learning has not been sufficiently addressed for learning action representations (16 out of 152). Observe that this lack of exploratory learning immediately relates to the maturity of a model by means of whether a representation is evaluated on a real robot or not. Clearly, learning and evaluating action representations on real robotic platforms strengthens the maturity of a representation.

One of the hallmark features of the human mind is its robustness to noisy or corrupt sensory inputs. This capacity stems for one central feat of human development: lifelong learning in a noisy and dynamic environment. Hence, only by grounding observations in real-world experiences, our minds are able to develop robust motor control (Harnad 1990). It is thence evident that for action representations in robotics we conjecture that such robustness yielding from grounding experiences in real-world observations is paramount. Besides, the capacity to generalize to new situations also plays a major role when it comes to robustness. Obviously, not being able to generalize to novel situations likely indicates a very weak model. Looking at Figure 3 we see that nearly all categorized methods generalize to novel situations (146 out of 152) indicating high robustness of most approaches. Yet, looking at how many of those ground effects shows quite a different picture. Not even a third of those (46 out of 146) actually ground effects by real-world experiences, hence now undermining the robustness of the remaining approaches. Correlating

these numbers with the means of exploitation however immediately reveals that 73 of the models not grounding effects are exploited only for recognition (observe that the remaining three recognition models do ground effects). Undoubtedly, recognition is feasible without grounding effects. For the remaining 27 models we unfortunately either lack the relevant data, or, in the other case, these models mostly do single-/multistep prediction using models trained by video sequences. The above epitomizes again the prevalence of recognition models which just do not require effect grounding. In the remaining cases, we conjecture that this due to a neglect of selective attention (only 27 out of 152 do so). Naturally, selective attention allows the curse of dimensionality to be tackled by focusing only on the stimuli that are relevant, thereby catalyzing the grounding of effects. Figure 4 however reveals that only 11 out of those 27 models ground effects. We claim that future action representations need to capitalize on selective attention for facilitating effect grounding thus drastically improving robustness.

Compiling the above, usability is essential for action representations. Current issues as discussed however could be tackled by implementing and especially evaluating a representation directly on a real robotic platform. Such an approach immediately unlocks the grounding of effects and consequently strengthen the maturity of the evaluated representation. By additionally considering selective attention one readily ends up with a representation substantially more robust than most current approaches.

5.5 A Few Last Words on Action and Activity Recognition Datasets

Inspired by a recent survey of Chaquet et al. (2013) we also investigated the use and wide-spread uptake of datasets as reported by the categorized papers. Table 1 shows the resulting distribution of datasets as reported by our classification. In total, 41 different datasets have been used by various papers if evaluating an action representation using a benchmark (80 out of 152, see Figure 3). Investigating the actual usage count of the various datasets, Table 1 shows a similar preference pattern as Table 5 of Chaquet et al.'s (2013) survey. For example, KTH, Weizmann and IXMAS are all among the top five datasets used. If learning of action representations is possible from datasets for action recognition, evaluating the relevance of the representation for robotics should be similarly straightforward (*c.f.* computer vision (Wu et al. 2015; Russakovsky et al. 2015)). It is thus critical to define suitable, standardized datasets to learn action knowledge and corresponding benchmarking setups to properly evaluate the representation. This would greatly enhance quantitative comparison of different approaches, simply because the baseline is the same.

A more severe usage pattern is shown by Table 2 in that only a small fraction of papers evaluated on benchmark datasets used more than two datasets. Evaluating a model only on one or two datasets may drastically falsify results regarding generalization capabilities, simply because of focusing only on a small set of actions captured in just a couple of environments. Considering multiple datasets for evaluation—in line with the above—further allows for more insight into the behavior and capabilities of a model,

and therefore for more robust models by virtue of better understanding.

We submit that applying more diversity in evaluating models on benchmarks, that is, using multiple and especially commonly used datasets, would greatly advance research on action representations in robotics. This advancement eventually capitalizes on deeper insight and understanding of how these various models actually achieve their desired outcome by meaningful quantitative comparisons.

Dataset	# Usage
KTH (Schüldt et al. 2004)	15
Weizmann (Blank et al. 2005)	13
IXMAS (Weinland et al. 2006)	8
MSR-Action-3D (Li et al. 2010)	7
HMDB (Kuehne et al. 2011)	4
3D Action Pairs (Oreifej and Liu 2013)	2
50 Salads (Stein and McKenna 2013)	2
ADLs (Pirsiavash and Ramanan 2012)	2
CAD-60 (Sung et al. 2012)	2
CMU-MoCap (CMU 2003)	2
Florence3D Actions (Seidenari et al. 2013)	2
HDM05 (Müller et al. 2007)	2
Hollywood2 (Marszalek et al. 2009)	2
MoPrim (Reng et al. 2005)	2
MSR-II (Cao et al. 2010)	2
MSR Daily Activiy (Wang et al. 2012)	2
UTKinect-Action (Xia et al. 2012)	2
YouTube (Liu et al. 2009)	2
UCF-101 (Soomro et al. 2012)	2
UCF-Sports (Rodriguez et al. 2008)	2
Berkeley-MHAD (Ofli et al. 2013)	1
ChaLearn Gesture (Guyon et al. 2012)	1
CHEMLAB corpus (Vitkute-Adzgauskiene et al. 2014)	1
FBG (Hwang et al. 2007)	1
Fish-action (Rahman et al. 2012)	1
G3D (Bloom et al. 2012)	1
Human Grasp (Schenatti et al. 2003)	1
JIGSAWS (Gao et al. 2014)	1
ManiAc (Aksoy et al. 2015)	1
MSRC-12 (Fothergill et al. 2012)	1
MuHAVi (Singh et al. 2010)	1
Olympic-Sports (Niebles et al. 2010)	1
Ravel (Alameda-Pineda et al. 2011)	1
RGBD-HUDAACT (Ni et al. 2013)	1
Reading Act (Chen et al. 2014)	1
Robust (Gorelick et al. 2007)	1
Stanford-40 Actions (Yao et al. 2011)	1
SYSU-3D-HOI (Science and Lab 2017)	1
TACoS (Regneri et al. 2013)	1
UMD (Veeraraghavan et al. 2006)	1
UT-Interaction (Ryoo and Aggarwal 2010)	1
YouTube (Liu et al. 2009)	1

Table 1. Datasets used for benchmarking in various categorized papers with respective usage count.

# Datasets	# Papers
4	5
3	7
2	13
1	31

Table 2. Total number of datasets used by various categorized papers.

6   Open research challenges

Our classification and the resulting discussion from the previous section show that action representations in robotics have been intensively studied in recent years. However, our discussions from Sections 5.1–5.5 also reveal that the current state of the art regarding action representations in robotics is still in an early stage and currently suffers from multiple issues. Below we provide an overview of the central research challenges as revealed by the results of our analysis. We believe that addressing these is paramount to successfully advance research on action representations in robotics.

- *Intensifying effect-centricity and effect grounding* Grounding of effects in real-world percepts is one of the key challenges from our point of view. Clearly, due to the vast amount if information available at each moment from both the self and the environment this is a hard challenge. Yet, doing so is critical to improve the quality of a model. As mentioned below, selective attention is one of the keys in handling this vast amount of data. Yet, we further claim that the capability of processing multi-modal percepts also substantially catalyzes the grounding of effects.
- *Coupled Forward and inverse models* One of the central advantages of biomimetic models, especially in the field of motor control and thence action representations, is their postulation of the need for inverse models. It is thence necessary to carefully reconsider current results in neuroscience and motor cognition (*c.f.* Section 2) to tackle the prevalent lack of inverse models. Doing so, among other benefits, readily unlocks the capacity of bidirectional effect associativity as well as performing motor imagery (Jeannerod 2006).
- *Exploiting language for action understanding* The compositional and semantically-rich nature of language is a strong prior for action understanding. Language provides precise and unambiguous semantics when it comes to describing actions. Therefore, we claim that besides grounding of effects in real-world observations, rooting the meaning of an action in natural language further boosts both learning and properly understanding an action. In the long run, this allows learning of more abstract, i.e., disembodied, and thence useful action representations.
- *Intrinsically-motivated, exploratory, semi- and self-supervised learning* Importantly, humans learn by observation and subsequent exploration and interaction with their environment. Following this central motive, it is crucial to allow computational agents to learn relevant concepts with minimal prior information. This allows for progressive learning of representations of the external world as well as of the self. Clearly, this requires an agent to be accordingly motivated as well as the capacity of self-supervising its learning efforts. This ultimately culminates in using already-learned concepts, to both drive and supervise the learning autonomously. We claim that learning in such a way yields stronger autonomy compared to classic supervised learning and thence merits more attention.
- *Selective attention* Again, we argue similarly to Zech et al. (2017) that selective attention is an important aspect for focused perception by blocking out clutter and noise. Contrary to our reasoning in the case of affordance however, here we claim that selective attention should be ascribed a central role as a precursor for grounding effects by successfully tackling the curse of dimensionality by only considering those stimuli which are relevant for grounding the observed effects, thus drastically boosting the robustness of different representations. Observe the immediate complementarity to the above challenge regarding effect centricity and grounding of effects.
- *Solving the correspondence problem* Similarly to Zech et al. (2017) we claim here that it is of utmost importance to solve the correspondence problem in robotics, i.e., mapping of observed motions. This would address current drawbacks in both learning from demonstration and in understanding actions from an observer's point of view. Especially in the event of action representations this would allow closing the gap from recognition to re-execution. Observe that this also requires intensified research towards constructing internal models of the agent's self.
- *Sequence-based modeling* The capability of composing compound actions, e.g., pick-and-place, out of more granular, atomic actions is a central capacity of mammalian motor control. Our minds do not store complete motor programs for each and every action but rather dynamically synthesize them out of more general building blocks for seamless action execution (*c.f.* Section 2). Clearly, such a capacity is also paramount for action representations in robotics especially with regards to generalizability but also scalability at a computational level.

Observe that there exists a substantial intersection of the above challenges with those identified by Zech et al. (2017) in the case of affordance research in robotics. This however is not surprising given the strong relation between actions and affordances, the latter being a key driver in action selection. This intersection clearly resembles the strong interrelation of these two complementary fields of research and thus motivates joint research efforts.

7   Conclusion

Action representations are a key ingredient of autonomy in robots. In this article we thus made three major contributions relevant for this field of research. After a thorough survey of the meaning of action as well as contemporary definitions and opinions from various associated scientific disciplines we ended with a seminal definition of action relevant to robotics (*c.f.* Section 2). This treatise thence paved the way for the first major contribution of our article, a taxonomy of action representations in robotics (*c.f.* Section 3). This allowed us to conduct our second major contribution, a systematic review of existing work on action representations in robotics. Identified publications subsequently were categorized using our taxonomy, yielding the results for our third contribution in the form of

an in-depth discussion of existing research on action representations in robotics (*c.f.* Section 5). This discussion finally culminated in the identification of key research challenges we deem fundamental for advancing research on action representations in robotics (*c.f.* Section 6).

Summarizing our work we report that for now one of the central drawbacks in action research in robotics is the crucial lack of a common notion of both action and action representation in robotics. However, this shall not raise the impression that current state of the art work is useless. On the contrary, existing results act both as a foundation and guidance towards how to advance action research in robotics. Accordingly, in Section 6 we identified future courses of actions for action research in robotics. We believe that intensifying research in these fields prolifically unlocks novel motor-cognitive capabilities in autonomous agents towards both more autonomy and dexterity.

A Abbreviations for Classification

Tables 3 and 4 show the various abbreviations as used in the classification depicted in Tables **??** and **??**.

Abbreviation	Definition
b	both
n	No
ns	not specified
y	Yes
Per	**Perspective**
li	limb
ag	agent
ob	observer
St	**Stimuli**
e	exteroceptive
p	proprioceptive
SA	**Selective Attention**
Grn	**Granularity**
lo	Local
me	Meso
gl	Global
Abs	**Abstraction**
a	atomic
c	compound
Com	**Competition**
Seq	**Sequencing**
Mot	**Motivation**
in	intrinsic
ex	extrinsic
Acq	**Acquisition**
com	combination
d	demonstration
exp	exploration
gt	ground truth
hc	hard coded
l	language
Pred	**Prediction**
cla	classification

inf	inference
opt	optimization
reg	regression
Exp	**Exploitation**
ep	effect prediction
l	language
p	planning
r	recognition
sa	self-assessment
sp	single-/multi-step prediction
Lrn	**Learning**
off	offline
on	online
Disc	**Discretization**
ca	categorical
co	continuous
Gnd	**Grounding**
Asso	**Associativity**
ud	unidirectional
bd	bidirectional
Corr	**Effect Correspondence**
by	body
env	environment

Table 3. Abbreviations for action model.

Abbreviation	Definition
Form	**Formulation**
MAT	mathematical
BIO	biomimetic
Train	**Training**
S	supervised
SELF	self-supervised
SEMI	semi-supervised
U	unsupervised
Eval	**Evaluation**
BM	benchmark
RR	real robot
SIM	simulation
VR	virtual reality
C	combination

Table 4. Abbreviations for computational model.

B Abbreviations for Methods and Features

Table 5 lists the definitions of abbreviations denoting the various features and methods as reported by the papers categorized in Tables **??** and **??**.

Abbreviation	Definition
*MB	Model-Based
*MF	Model-Free
AE	Autoencoder
ANN	Artificial Neural Network
ASOM	Associative Som
BN	Bayesian Network
BOF	Bag-Of-Features
BOO	Bag-Of-Objects
BOW	Bag-Of-Words
BP	Belief Propagation
CMAC	Cerebellar Model Articulation Controller
CNN	Convolutional Neural Network

CRF	Conditional Random Field
CS	Conceptual Spaces
CTRNN	Continuous Time RNN
DAG-RNN	Directed Acyclic Graph RNN
DBN	Dynamic Bayesian Networks
DCNN	Deep CNN
DMP	Dynamic Movement Primitive
DNN	Deep Neural Network
DP	Dynamic Programming
DS	Dynamical System
DTW	Dynamic Time Warping
ECV	Early Cognitive Vision
EDM	Euclidean Distance Matrix
EKF	Extended Kalman Filter
ELM	Extreme Learning Machine
EM	Expectation-Maximization
EMG	Electromyography
FFT	Fast-Fourier Transform
FREAK	Fast Retina Keypoint
FSM	Finite-state Machine
FSTM	Feasible Situation Transition Manifold
GLOH	Gradient Location and Orientation Histogram
GMM	Gaussian Mixture Model
GMR	Gaussian Mixture Regression
GP	Gaussian Process
GPR	Gaussian Process Regression
GWR	Growing When Required Network
HHMM	Hierarchical Hidden Markov Model
HMM	Hidden Markov Model
HOG	History Of Gradients
HOS	Histogram Of Silhouette
HPNNA	Hierarchical Programmable NN Architecture
ICA	Independent Component Analysis
IMU	Internal Measurement Unit
KDE	Kernel Density Estimation
KD-Tree	k-Dimensional Tree
k-NN	k-Nearest Neighbor
LCSS	Longest Common Subsequence
LD	Levenshtein Distance
LSM	Liquid State Machine
LSTM	Long-Short Term Memory
LVQ	Learning Vector Quantization
MCSVM	Multiclass SVM
MDN	Mixture Density Network
MDP	Markov Decision Process
MKL	Multiple Kernel Learning
MMI	Maximization of Mutual Information
MMM	Master Motor Map
MMR	Maximum Margin Regression
MNN	Modular Neural Networks
MoFREAK	Motion-Based FREAK
MSER	Maximally Stable Extremal Regions
MTRNN	Multiple Timescales RNN
NBNN	Naive Bayes Nearest Neighbor
NF	Neural Field
NGLD	Normalized Google-Like Distance
NLP	Natural Language Processing
NMF	Negative Matrix Factorization
NNC	Nearest Neighbour Classifier

NNMF	Non-NMF
NN	Neural Network
PCA	Principal Component Analysis
PCA-STOP	PCA Space-Time Occupancy Patterns
PDI	Positional Distribution Information
PHMM	Parametric HMM
PLSA	Probabilistic Latent Semantic Analysis
PMP	Passive Motion Paradigm
PMT	Projected Motion Template
PP	Purr-Puss
PSVM	Probabilistic SVM
PVS	Predicate Vector Sequence
QTC	Qualitative Trajectory Calculus
RBF	Radial Basis Function
RF	Random Forest
RGB-D	Red-Green-Blue-Depth
RGB	Red-Green-Blue
RL	Reinforcement Learning
RNNPB	RNN with Parametric Bias
RNN	Recurrent Neural Network
SCFG	Stochastic Context Free Grammar
SEC	Semantic Event Chain
SFA	Slow Feature Analysis
SIFT	Scale-Invariant Feature Transform
SOM	Self Organizing Map
SPHOF	Spatial Pyramid Histogram of Optical Flow
SSM	Self-Similarity Matrix
SSP	Space Salient Pairwise Feature
STDP	Spike-Timing Dependent Plasticity
STIP	Spatio-temporal interest points
STV	Spatio-Temporal Volumes
SVM	Support Vector Machine
SVR	Support Vector Regression
TSP	Time Salient Pairwise Feature
VAE	Variational AE
VMT	Volume Motion Template
WSM	Word Space Model

Table 5. Abbreviations for methods an features

C Classification of selected publications

Tables **??** and **??** show the full classification of all selected publications. These results are also available online at https://iis.uibk.ac.at/public/survey/ActionRepresentation/.

Acknowledgments

References

Acosta Calderon CA, Mohan RE and Zhou C (2010) Teaching new tricks to a robot learning to solve a task by imitation. In: *2010 IEEE Conference on Robotics, Automation and Mechatronics*. IEEE. DOI:10.1109/ramech.2010.5513180.

Aein MJ, Aksoy EE, Tamosiunaite M, Papon J, Ude A and Worgotter F (2013) Toward a library of manipulation actions based on semantic object-action relations. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. DOI:10.1109/iros.2013.6697011.

Ahad MAR, Tan J, Kim H and Ishikawa S (2010) Action recognition by employing combined directional motion history and energy images. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. IEEE. DOI:10.1109/cvprw.2010.5543160.

Ahmad M and Lee SW (2010) Variable silhouette energy image representations for recognizing human actions. *Image and Vision Computing* 28(5): 814–824. DOI:10.1016/j.imavis.2009.09.018.

Ahmadzadeh SR, Paikan A, Mastrogiovanni F, Natale L, Kormushev P and Caldwell DG (2015) Learning symbolic representations of actions from human demonstrations. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. DOI:10.1109/icra.2015.7139728.

Aksoy EE, Orhan A and Wörgötter F (2016a) Semantic decomposition and recognition of long and complex manipulation action sequences. *International Journal of Computer Vision* 122(1): 84–115. DOI:10.1007/s11263-016-0956-8.

Aksoy EE, Tamosiunaite M, Vuga R, Ude A, Geib C, Steedman M and Worgotter F (2013) Structural bootstrapping at the sensorimotor level for the fast acquisition of action knowledge for cognitive robots. In: *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE. DOI:10.1109/devlrn.2013.6652537.

Aksoy EE, Tamosiunaite M and Wörgötter F (2015) Model-free incremental learning of the semantics of manipulation actions. *Robotics and Autonomous Systems* 71: 118–133.

Aksoy EE, Zhou Y, Wachter M and Asfour T (2016b) Enriched manipulation action semantics for robot execution of time constrained tasks. In: *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE. DOI:10.1109/humanoids.2016.7803262.

Alameda-Pineda X, Sanchez-Riera J, Franch V, Wienke J, Cech J, Kulkarni K, Deleforge A and Horaud R (2011) *The RAVEL data set*. PhD Thesis, INRIA.

Altahhan A (2015) Deep feature-action processing with mixture of updates. In: *Neural Information Processing*. Springer International Publishing, pp. 1–10. DOI:10.1007/978-3-319-26561-2_1.

Andry P, Gaussier P, Nadel J and Hirsbrunner B (2004) Learning invariant sensorimotor behaviors: A developmental approach to imitation mechanisms. *Adaptive Behavior* 12(2): 117–140. DOI:10.1177/105971230401200203.

Aristotle (1934) *Nicomachean Ethics*. Loeb Classical Library. Cambridge, MA: Harvard University Press.

Asfour T, Welke K, Ude A, Azad P and Dillmann R (2008) Perceiving objects and movements to generate actions on a humanoid robot. In: *Lecture Notes in Electrical Engineering*. Springer US, pp. 41–55. DOI:10.1007/978-0-387-75523-6_4.

Babič J, Hale JG and Oztop E (2011) Human sensorimotor learning for humanoid robot skill synthesis. *Adaptive Behavior* 19(4): 250–263. DOI:10.1177/1059712311411112.

Bamert L and Mast FW (2009) *Action Representation*. Springer Berlin Heidelberg, pp. 32–34.

Bartels G, Kresse I and Beetz M (2013) Constraint-based movement representation grounded in geometric features. In: *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE. DOI:10.1109/humanoids.2013.7030027.

Bernstein NA (1996) *On Dexterity and its development*. Lawrence Erlbaum Associates.

Bessiere P, Dedieu E and Mazer E (1994) Representing robot/environment interactions using probabilities: the "beam in the bin" experiment. In: *Proceedings of PerAc '94. From Perception to Action*. IEEE Comput. Soc. Press. DOI:10.1109/fpa.1994.636093.

Bhat AA and Mohan V (2015) How iCub learns to imitate use of a tool quickly by recycling the past knowledge learnt during drawing. In: *Biomimetic and Biohybrid Systems*. Springer International Publishing, pp. 339–347. DOI:10.1007/978-3-319-22979-9_33.

Blank M, Gorelick L, Shechtman E, Irani M and Basri R (2005) Actions as space-time shapes. In: *The Tenth IEEE International Conference on Computer Vision (ICCV'05)*. pp. 1395–1402. URL https://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html.

Bloom V, Makris D and Argyriou V (2012) G3D: A Gaming Action Dataset and Real-time Action Recognition Evaluation Framework. In: *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. pp. 7–12.

Botvinick MM (2008) Hierarchical models of behavior and prefrontal function. *Trends in cognitive sciences* 12 5: 201–8.

Buonamente M, Dindo H and Johnsson M (2013) Recognizing actions with the associative self-organizing map. In: *2013 XXIV International Conference on Information, Communication and Automation Technologies (ICAT)*. IEEE. DOI:10.1109/icat.2013.6684076.

Cantrell R, Schermerhorn P and Scheutz M (2011) Learning actions from human-robot dialogues. In: *2011 RO-MAN*. IEEE. DOI:10.1109/roman.2011.6005199.

Cao L, Liu Z and Huang TS (2010) Cross-dataset action detection. In: *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*. IEEE, pp. 1998–2005. URL http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/default.htm.

Chaaraoui AA, Climent-Pérez P and Flórez-Revuelta F (2012) An efficient approach for multi-view human action recognition based on bag-of-key-poses. In: *Human Behavior Understanding*. Springer Berlin Heidelberg, pp. 29–40. DOI:10.1007/978-3-642-34014-7_3.

Chaquet JM, Carmona EJ and Fernández-Caballero A (2013) A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding*

117(6): 633–659.

Chella A, Frixione M and Gaglio S (2000) Towards a conceptual representation of actions. Springer Berlin Heidelberg, pp. 333–344. DOI:10.1007/3-540-46238-4_29.

Chen L, Wei H and Ferryman J (2014) ReadingAct RGB-d action dataset and human action recognition from local features. *Pattern Recognition Letters* 50: 159–169. DOI:10.1016/j.patrec.2013.09.004.

Chuang LW, Lin CY and Cangelosi A (2012) Learning of composite actions and visual categories via grounded linguistic instructions: Humanoid robot simulations. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE. DOI:10.1109/ijcnn.2012.6252520.

Cisek P and Kalaska JF (2010) Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience* 33(1): 269–298.

Claßen J, Röger G, Lakemeyer G and Nebel B (2011) Platas—integrating planning and the action language golog. *KI - Künstliche Intelligenz* 26(1): 61–67. DOI:10.1007/s13218-011-0155-2.

CMU (2003) Graphics lab motion capture. URL http://mocap.cs.cmu.edu/. Contact: jkh+mocap@cs.cmu.edu.

Cooper RP and Shallice T (2006) Hierarchical schemas and goals in the control of sequential behavior. .

Davidson D (2001) *Essays on Actions and Events: Philosophical Essays*. Clarendon Press.

De Kleijn R, Kachergis G and Hommel B (2014) Everyday robotic action: Lessons from human action control. *Frontiers in Neurorobotics* 8: 1–9. ECollection.

Desmurget M and Grafton S (2000) Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences* 4(11): 423 – 431. DOI:https://doi.org/10.1016/S1364-6613(00)01537-0.

Dindo H and Chella A (2013) What will you do next? a cognitive model for understanding others' intentions based on shared representations. In: *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*. Springer Berlin Heidelberg, pp. 253–266. DOI:10.1007/978-3-642-39405-8_29.

Dindo H, Presti LL, Cascia ML, Chella A and Dedić R (2017) Hankelet-based action classification for motor intention recognition. *Robotics and Autonomous Systems* 94: 120–133. DOI:10.1016/j.robot.2017.04.003.

Do M, Schill J, Ernesti J and Asfour T (2014) Learn to wipe: A case study of structural bootstrapping from sensorimotor experience. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. DOI:10.1109/icra.2014.6907103.

Donnarumma F, Prevete R, de Giorgio A, Montone G and Pezzulo G (2015) Learning programs is better than learning dynamics: A programmable neural network hierarchical architecture in a multi-task scenario. *Adaptive Behavior* 24(1): 27–51. DOI:10.1177/1059712315609412.

Droniou A, Ivaldi S and Sigaud O (2014) Learning a repertoire of actions with deep neural networks. In: *4th International Conference on Development and Learning and on Epigenetic Robotics*. IEEE. DOI:10.1109/devlrn.2014.6982986.

Dum R and Strick P (1991) The origin of corticospinal projections from the premotor areas in the frontal lobe. *Journal of Neuroscience* 11(3): 667–689. DOI:10.1523/JNEUROSCI.

11-03-00667.1991.

Dum RP and Strick PL (1996) Spinal cord terminations of the medial wall motor areas in macaque monkeys. *Journal of Neuroscience* 16(20): 6513–6525. DOI:10.1523/JNEUROSCI.16-20-06513.1996.

Elsner B and Hommel B (2001) Effect Anticipation and Action Control. *Journal of experimental psychology: human perception and performance* 27(1): 229.

Endres D, Chiovetto E and Giese MA (2015) Bayesian approaches for learning of primitive-based compact representations of complex human activities. In: *Dance Notations and Robot Motion*. Springer International Publishing, pp. 117–137. DOI:10.1007/978-3-319-25739-6_6.

Englert P, Paraschos A, Deisenroth MP and Peters J (2013) Probabilistic model-based imitation learning. *Adaptive Behavior* 21(5): 388–403. DOI:10.1177/1059712313491614.

Farhadi A and Tabrizi MK (2008) Learning to recognize activities from the wrong view point. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 154–166. DOI:10.1007/978-3-540-88682-2_13.

Fihl P, Holte MB, Moeslund TB and Reng L (2006) Action recognition using motion primitives and probabilistic edit distance. In: *Articulated Motion and Deformable Objects*. Springer Berlin Heidelberg, pp. 375–384. DOI:10.1007/11789239_39.

Fothergill S, Mentis H, Kohli P and Nowozin S (2012) Instructing people for training gestural interactive systems. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 1737–1746.

Francis BA and Wonham WM (1976) The Internal Model Principle of Control Theory. *Automatica* 12(5): 457–465.

Fuster J (1999) *Memory in the Cerebral Cortex (Second Edition)*. ISBN 0-262-56124-7 (pb).

Gao Y, Vedula SS, Reiley CE, Ahmidi N, Varadarajan B, Lin HC, Tao L, Zappella L, Béjar B, Yuh DD et al. (2014) Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling. In: *MICCAI Workshop: M2CAI*, volume 3. p. 3.

Gibson JJ (1966) *The Senses Considered as Perceptual Systems*. Houghton Mifflin.

Gibson JJ (1979) *The Ecological Approach to Visual Perception*. Psychology Press.

Gorelick L, Blank M, Shechtman E, Irani M and Basri R (2007) Actions as space-time shapes. *IEEE transactions on pattern analysis and machine intelligence* 29(12): 2247–2253.

Grafton S, Aziz-Zadeh L and Ivry R (2009) Relative hierarchies and the representation of action : 641–652.

Grave K and Behnke S (2012) Incremental action recognition and generalizing motion generation based on goal-directed features. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. DOI:10.1109/iros.2012.6386116.

Grinke E, Tetzlaff C, Wörgötter F and Manoonpong P (2015) Synaptic plasticity in a recurrent neural network for versatile and adaptive behaviors of a walking robot. *Frontiers in Neurorobotics* 9. DOI:10.3389/fnbot.2015.00011.

Gritai A, Sheikh Y, Rao C and Shah M (2009) Matching trajectories of anatomical landmarks under viewpoint, anthropometric and temporal transforms. *International Journal of Computer Vision*

84(3): 325–343. DOI:10.1007/s11263-009-0239-8.

Gross R (2015) *Psychology: The Science of Mind and Behaviour.* Hodder Education.

Guerra-Filho G and Aloimonos Y (2007) A Language for Human Action. *Computer* 40(5): 42–51.

Guha A, Yang Y, Fermuuller C and Aloimonos Y (2013) Minimalist plans for interpreting manipulation actions. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems.* IEEE. DOI:10.1109/iros.2013.6697213.

Guyon I, Athitsos V, Jangyodsuk P, Hamner B and Escalante HJ (2012) Chalearn gesture challenge: Design and first results. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on.* IEEE, pp. 1–6.

Hajimirsadeghi H, Ahmadabadi MN and Araabi BN (2013) Conceptual imitation learning based on perceptual and functional characteristics of action. *IEEE Transactions on Autonomous Mental Development* 5(4): 311–325. DOI:10. 1109/tamd.2013.2263833.

Hamilton A and Grafton ST (2007) The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor foundations of higher cognition* 22: 381–408.

Hamilton AFdC and Grafton ST (2006) Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience* 26(4): 1133–1137.

Hamilton AFdC and Grafton ST (2008) Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex* 18(5): 1160–1168. DOI:10.1093/cercor/bhm150. URL http://dx.doi.org/10.1093/cercor/bhm150.

Haneda A, Okada K and Inaba M (2008) Realtime manipulation planning system integrating symbolic and geometric planning under interactive dynamics siumlator. In: *2008 IEEE International Conference on Mechatronics and Automation.* IEEE. DOI:10.1109/icma.2008.4798893.

Hardwick RM, Caspers S, Eickhoff SB and Swinnen SP (2017) Neural Correlates of Motor Imagery, Action Observation, and Movement Execution: A Comparison Across Quantitative Meta-Analyses. *bioRxiv* DOI:10.1101/198432.

Harnad S (1990) The symbol grounding problem. *Physica D: Nonlinear Phenomena* 42(1): 335 – 346.

Hasson U, Chen J and Honey CJ (2015) Hierarchical Process Memory: Memory as an Integral Component of Information Processing. *Trends in cognitive sciences* 19(6): 304–313.

Herzog DL and Krüger V (2012) Tracking in action space. In: *Trends and Topics in Computer Vision.* Springer Berlin Heidelberg, pp. 100–113. DOI:10.1007/978-3-642-35749-7_8.

Hobaiter C (2017) What is a gesture? A systematic approach to defining gestural repertoires. *Neuroscience and Biobehavioral Reviews* 82: 3–12.

Hofer S and Brock O (2016) Coupled learning of action parameters and forward models for manipulation. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* IEEE. DOI:10.1109/iros.2016.7759573.

Hongeng S and Wyatt J (2008) Learning causality and intentional actions. In: *Towards Affordance-Based Robot Control.* Springer Berlin Heidelberg, pp. 27–46. DOI:10.1007/978-3-540-77915-5_3.

Hourdakis E and Trahanias P (2012) Computational modeling of observational learning inspired by the cortical underpinnings of human primates. *Adaptive Behavior* 20(4): 237–256. DOI: 10.1177/1059712312445902.

Hwang BW, Kim S and Lee SW (2007) A full-body gesture database for human gesture analysis. *International Journal of Pattern Recognition and Artificial Intelligence* 21(06): 1069–1084.

Hwang J and Tani J (2018) Seamless integration and coordination of cognitive skills in humanoid robots: A deep learning approach. *IEEE Transactions on Cognitive and Developmental Systems* 10(2): 345–358. DOI:10.1109/tcds.2017.2714170.

Ijjina EP and Krishna Mohan C (2016) Classification of human actions using pose-based features and stacked auto encoder. *Pattern Recognition Letters* 83: 268–277. DOI:10.1016/j. patrec.2016.03.021.

Jeannerod M (1984) The timing of natural prehension movements. *Journal of Motor Behavior* 16(3): 235–254. DOI:10.1080/ 00222895.1984.10735319. PMID: 15151851.

Jeannerod M (1986) The formation of finger grip during prehension. a cortically mediated visuomotor pattern. *Behavioural Brain Research* 19(2): 99 – 116. DOI:https://doi.org/10.1016/ 0166-4328(86)90008-2.

Jeannerod M (2006) *Motor Cognition: What Actions tell the Self.* 42. Oxford University Press.

Jeon Y, Sandhan T and Choi JY (2015) Robust feature extraction for shift and direction invariant action recognition. In: *Lecture Notes in Computer Science.* Springer International Publishing, pp. 321–329. DOI:10.1007/978-3-319-24078-7_32.

Ji X and Liu H (2009) View-invariant human action recognition using exemplar-based hidden markov models. In: *Intelligent Robotics and Applications.* Springer Berlin Heidelberg, pp. 78–89. DOI:10.1007/978-3-642-10817-4_8.

Ji X, Liu H and Li Y (2010) A new framework for view-invariant human action recognition. In: *Advanced Information and Knowledge Processing.* Springer London, pp. 71–93. DOI: 10.1007/978-1-84996-329-9_4.

Ji XF, Wu QQ, Ju ZJ and Wang YY (2014) Study of human action recognition based on improved spatio-temporal features. *International Journal of Automation and Computing* 11(5): 500–509. DOI:10.1007/s11633-014-0831-4.

Ji Y, Yang Y, Xu X and Shen HT (2018) One-shot learning based pattern transition map for action early recognition. *Signal Processing* 143: 364–370. DOI:10.1016/j.sigpro.2017.06.001.

Jiang H and Martin DR (2008) Finding actions using shape flows. In: *Lecture Notes in Computer Science.* Springer Berlin Heidelberg, pp. 278–292. DOI:10.1007/978-3-540-88688-4_21.

Johnson SH and Grafton ST (2003) From 'acting on' to 'acting with': the functional anatomy of object-oriented action schemata. In: *Neural Control of Space Coding and Action Production, Progress in Brain Research*, volume 142. Elsevier, pp. 127 – 139. DOI:https://doi.org/10.1016/S0079-6123(03) 42010-4. URL http://www.sciencedirect.com/ science/article/pii/S0079612303420104.

Junejo IN, Dexter E, Laptev I and Pérez P (2008) Cross-view action recognition from temporal self-similarities. In: *Lecture Notes in Computer Science.* Springer Berlin Heidelberg, pp. 293–306. DOI:10.1007/978-3-540-88688-4_22.

Kaiser M (1997) Transfer of elementary skills via human-robot interaction. *Adaptive Behavior* 5(3-4): 249–280. DOI:10.1177/105971239700500303.

Karn NK and Jiang F (2016) Improved GLOH approach for one-shot learning human gesture recognition. In: *Biometric Recognition*. Springer International Publishing, pp. 441–452. DOI:10.1007/978-3-319-46654-5_49.

Keele SW and Jennings PJ (1992) Attention in the representation of sequence: Experiment and theory. *Human Movement Science* 11(1): 125 – 138. DOI:https://doi.org/10.1016/0167-9457(92)90055-G.

Kemke C (2006) Natural language communication between human and artificial agents. In: *Agent Computing and Multi-Agent Systems*. Springer Berlin Heidelberg, pp. 84–93. DOI:10.1007/11802372_11.

Kitchenham B (2004) Procedures for performing systematic reviews. *Keele, UK, Keele University* 33(2004): 1–26.

Kjellström H, Romero J and Kragić D (2011) Visual object-action recognition: Inferring object affordances from human demonstration. *Computer Vision and Image Understanding* 115(1): 81–90. DOI:10.1016/j.cviu.2010.08.002.

Koniusz P, Cherian A and Porikli F (2016) Tensor representations via kernel linearization for action recognition from 3d skeletons. In: *Computer Vision – ECCV 2016*. Springer International Publishing, pp. 37–53. DOI:10.1007/978-3-319-46493-0_3.

Krüger V (2006) Recognizing action primitives in complex actions using hidden markov models. In: *Advances in Visual Computing*. Springer Berlin Heidelberg, pp. 538–547. DOI:10.1007/11919476_54.

Krüger V and Grest D (2007) Using hidden markov models for recognizing action primitives in complex actions. In: *Image Analysis*. Springer Berlin Heidelberg, pp. 203–212. DOI:10.1007/978-3-540-73040-8_21.

Krüger V, Herzog DL, Baby S, Ude A and Kragić D (2010) Learning actions from observations. *IEEE Robotics & Automation Magazine* 17(2): 30–43. DOI:10.1109/mra.2010.936961.

Krüger V, Kragić D, Ude A and Geib C (2007) The meaning of action: a review on action recognition and mapping. *Advanced Robotics* 21(13): 1473–1501.

Krüger N, Geib C, Piater J, Petrick R, Steedman M, Wörgötter F, Ude A, Asfour T, Kraft D, Omrčen D, Agostini A and Dillmann R (2011) Object–action complexes: Grounded abstractions of sensory–motor processes. *Robotics and Autonomous Systems* 59(10): 740–757. DOI:10.1016/j.robot.2011.05.009.

Krüger V and Herzog D (2013) Tracking in object action space. *Computer Vision and Image Understanding* 117(7): 764–789. DOI:10.1016/j.cviu.2013.02.002.

Kuehne H, Jhuang H, Garrote E, Poggio T and Serre T (2011) Hmdb: a large video database for human motion recognition. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, pp. 2556–2563. URL http://serre-lab.clps.brown.edu/resource/HMDB/.

Kulkarni K, Boyer E, Horaud R and Kale A (2011) An unsupervised framework for action recognition using actemes. In: *Computer Vision – ACCV 2010*. Springer Berlin Heidelberg, pp. 592–605. DOI:10.1007/978-3-642-19282-1_47.

Kulkarni O, Parameswaran N and Nagarajan R (1989) Action representation for planning using truth maintenance system. In: *Fourth IEEE Region 10 International Conference TENCON*. IEEE. DOI:10.1109/tencon.1989.177098.

Kumar SH and Sivaprakash P (2013) New approach for action recognition using motion based features. In: *2013 IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES*. IEEE. DOI:10.1109/cict.2013.6558292.

Laaksonen J, Felip J, Morales A and Kyrki V (2010) Embodiment independent manipulation through action abstraction. In: *2010 IEEE International Conference on Robotics and Automation*. IEEE. DOI:10.1109/robot.2010.5509153.

Lallee (2010) Linking language with embodied and teleological representations of action for humanoid cognition. *Frontiers in Neurorobotics* DOI:10.3389/fnbot.2010.00008.

Layher G, Brosch T and Neumann H (2017) Real-time biologically inspired action recognition from key poses using a neuromorphic architecture. *Frontiers in Neurorobotics* 11. DOI:10.3389/fnbot.2017.00013.

Lea C, Vidal R and Hager GD (2016) Learning convolutional action primitives for fine-grained action recognition. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. DOI:10.1109/icra.2016.7487305.

Lee S and Chen J (1996) Robot skill discovery based on observed data(003) 5335512. In: *Proceedings of IEEE International Conference on Robotics and Automation*. IEEE. DOI:10.1109/robot.1996.506569.

Li W, Zhang Z and Liu Z (2010) Action recognition based on a bag of 3d points. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, pp. 9–14. URL http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/.

Liu J, Luo J and Shah M (2009) Recognizing realistic actions from videos "in the wild". In: *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on*. IEEE, pp. 1996–2003.

Liu M, Liu H, Sun Q, Zhang T and Ding R (2016) Salient pairwise spatio-temporal interest points for real-time activity recognition. *CAAI Transactions on Intelligence Technology* 1(1): 14–29. DOI:10.1016/j.trit.2016.03.001.

Mandik P (2005) Action-Oriented Representation. In: Brook A and Akins K (eds.) *Cognition and the Brain: The Philosophy and Neuroscience Movement*. Cambridge University Press, pp. 284–305.

Mansur A, Makihara Y and Yagi Y (2011) Action recognition using dynamics features. In: *2011 IEEE International Conference on Robotics and Automation*. IEEE. DOI:10.1109/icra.2011.5979900.

Markievicz I, Vitkute-Adzgauskiene D and Tamosiunaite M (2013) Semi-supervised learning of action ontology from domain-specific corpora. In: *Communications in Computer and Information Science*. Springer Berlin Heidelberg, pp. 173–185. DOI:10.1007/978-3-642-41947-8_16.

Marocco D (2010) Grounding action words in the sensorimotor interaction with the world: experiments with a simulated iCub humanoid robot. *Frontiers in Neurorobotics* DOI:10.3389/fnbot.2010.00007.

Marszalek M, Laptev I and Schmid C (2009) Actions in context. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, pp. 2929–2936. URL http://www.di.ens.fr/%7Elaptev/

`actions/hollywood2/`.

Martinet LE, Fouque B, Passot JB, Meyer JA and Arleo A (2008) Modelling the cortical columnar organisation for topological state-space representation, and action planning. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 137–147. DOI:10.1007/978-3-540-69134-1_14.

Maye A and Engel AK (2013) Extending sensorimotor contingency theory: prediction, planning, and action generation. *Adaptive Behavior* 21(6): 423–436. DOI:10.1177/1059712313497975.

Mikolajczyk K and Uemura H (2011) Action recognition with appearance–motion features and fast search trees. *Computer Vision and Image Understanding* 115(3): 426–438. DOI: 10.1016/j.cviu.2010.11.002.

Mohan V, Metta G, Zenzeri J and Morasso P (2010) Teaching humanoids to imitate 'shapes' of movements. In: *Artificial Neural Networks – ICANN 2010*. Springer Berlin Heidelberg, pp. 234–244. DOI:10.1007/978-3-642-15822-3_31.

Mohan V, Morasso P, Sandini G and Kasderidis S (2013) Inference through embodied simulation in cognitive robots. *Cognitive Computation* 5(3): 355–382. DOI:10.1007/s12559-013-9205-4.

Mugan J and Kuipers B (2012) Autonomous learning of high-level states and actions in continuous environments. *IEEE Transactions on Autonomous Mental Development* 4(1): 70–86. DOI:10.1109/tamd.2011.2160943.

Mukovskiy A, Vassallo C, Naveau M, Stasse O, Souères P and Giese MA (2017) Adaptive synthesis of dynamically feasible full-body movements for the humanoid robot HRP-2 by flexible combination of learned dynamic movement primitives. *Robotics and Autonomous Systems* 91: 270–283. DOI:10.1016/j.robot.2017.01.010.

Müller M, Röder T, Clausen M, Eberhardt B, Krüger B and Weber A (2007) Documentation mocap database hdm05. Technical Report CG-2007-2, Universität Bonn. URL `http://resources.mpi-inf.mpg.de/HDM05/`.

Mülling K, Kober J and Peters J (2011) A biomimetic approach to robot table tennis. *Adaptive Behavior* 19(5): 359–376. DOI: 10.1177/1059712311419378.

Naito E, Morita T and Amemiya K (2016) Body Representations in the Human Brain Revealed by Kinesthetic Illusions and their Essential Contributions to Motor Control and Corporeal Awareness. *Neuroscience Research* 104: 16 – 30.

Nakajo R, Murata S, Arie H and Ogata T (2015) Acquisition of viewpoint representation in imitative learning from own sensory-motor experiences. In: *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE. DOI:10.1109/devlrn.2015.7346166.

Natarajan P, Banerjee P, Khan FM and Nevatia R (2009) Graphical framework for action recognition using temporally dense STIPs. In: *2009 Workshop on Motion and Video Computing (WMVC)*. IEEE. DOI:10.1109/wmvc.2009.5399230.

Newton NW (2017) Understanding and Self-Organization. *Frontiers in Systems Neuroscience* 11(8): 1–9.

Ni B, Wang G and Moulin P (2013) Rgbd-hudaact: A color-depth video database for human daily activity recognition. In: *Consumer Depth Cameras for Computer Vision*. Springer, pp. 193–208.

Niebles JC, Chen CW and Fei-Fei L (2010) Modeling Temporal Structure of Decomposable Motion Segments for Activity Classification. In: Daniilidis K, Maragos P and Paragios N (eds.) *Computer Vision – ECCV 2010*. Springer Berlin Heidelberg, pp. 392–405.

Nishimoto R, Namikawa J and Tani J (2008) Learning multiple goal-directed actions through self-organization of a dynamic neural network model: A humanoid robot experiment. *Adaptive Behavior* 16(2-3): 166–181. DOI: 10.1177/1059712308089185.

Nishimoto R and Tani J (2009) Development process of functional hierarchy for actions and motor imagery. In: *2009 IEEE 8th International Conference on Development and Learning*. IEEE. DOI:10.1109/devlrn.2009.5175507.

Noda K, Kawamoto K, Hasuo T and Sabe K (2011) A generative model for developmental understanding of visuomotor experience. In: *2011 IEEE International Conference on Development and Learning (ICDL)*. IEEE. DOI:10.1109/devlrn.2011.6037357.

Nussbaum M (ed.) (1985) *Aristotle's De Motu Animalium: Text with Translation, Commentary, and Interpretive Essays*. Princeton Paperbacks. Princeton University Press.

Ofli F, Chaudhry R, Kurillo G, Vidal R and Bajcsy R (2013) Berkeley MHAD: A Comprehensive Multimodal Human Action Database. In: *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. pp. 53–60.

Ogawara K, Iba S, Tanuki T, Kimura H and Ikeuchi K (2001) Acquiring hand-action models by attention point analysis. In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*. IEEE. DOI: 10.1109/robot.2001.932594.

Ognibene D, Volpi NC, Pezzulo G and Baldassare G (2013a) Learning epistemic actions in model-free memory-free reinforcement learning: Experiments with a neuro-robotic model. In: *Biomimetic and Biohybrid Systems*. Springer Berlin Heidelberg, pp. 191–203. DOI:10.1007/978-3-642-39802-5_17.

Ognibene D, Wu Y, Lee K and Demiris Y (2013b) Hierarchies for embodied action perception. In: *Computational and Robotic Models of the Hierarchical Organization of Behavior*. Springer Berlin Heidelberg, pp. 81–98. DOI:10.1007/978-3-642-39875-9_5.

Oreifej O and Liu Z (2013) Hon4d: Histogram of oriented 4d normals for activity recognition from depth sequences. In: *Computer vision and pattern recognition (CVPR), 2013 IEEE conference on*. IEEE, pp. 716–723.

O'Shaughnessy B (1997) Trying (as the Mental 'Pineal Gland'). In: Mele AR (ed.) *The Philosophy of Action*. Oxford University Press, pp. 365–386.

Panchev C (2005) A spiking neural network model of multi-modal language processing of robot instructions. In: *Biomimetic Neural Learning for Intelligent Robots*. Springer Berlin Heidelberg, pp. 182–210. DOI:10.1007/11521082_11.

Panzner M and Cimiano P (2016) Comparing hidden markov models and long short term memory neural networks for learning action representations. In: *Lecture Notes in Computer Science*. Springer International Publishing, pp. 94–105. DOI: 10.1007/978-3-319-51469-7_8.

Parisi GI, Weber C and Wermter S (2015) Self-organizing neural integration of pose-motion features for human action recognition. *Frontiers in Neurorobotics* 9. DOI:10.3389/fnbot.2015.00003.

Park JC, Kim DS and Nagai Y (2017) Learning for goal-directed actions using RNNPB: Developmental change of "what to imitate". *IEEE Transactions on Cognitive and Developmental Systems* : 1–1DOI:10.1109/tcds.2017.2679765.

Patel M, Miro JV, Kragić D, Ek CH and Dissanayake G (2014) Learning object, grasping and manipulation activities using hierarchical HMMs. *Autonomous Robots* 37(3): 317–331. DOI:10.1007/s10514-014-9392-1.

Paulus M, van Dam W, Hunnius S, Lindemann O and Bekkering H (2011) Action-effect Binding by Observational Learning. *Psychonomic Bulletin & Review* 18(5): 1022.

Paxton C, Jonathan F, Kobilarov M and Hager GD (2016) Do what i want, not what i did: Imitation of skills by planning sequences of actions. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. DOI:10.1109/iros.2016.7759556.

Pazhoumand-Dar H, Lam CP and Masek M (2015) Joint movement similarities for robust 3d action recognition using skeletal data. *Journal of Visual Communication and Image Representation* 30: 10–21. DOI:10.1016/j.jvcir.2015.03.002.

Pezzulo G and Dindo H (2011) What should i do next? using shared representations to solve interaction problems. *Experimental Brain Research* 211(3-4): 613–630. DOI:10.1007/s00221-011-2712-1.

Pierobon M, Marcon M, Sarti A and Tubaro S (2005) Clustering of human actions using invariant body shape descriptor and dynamic time warping. In: *Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2005*. IEEE. DOI:10.1109/avss.2005.1577237.

Pieropan A, Salvi G, Pauwels K and Kjellstrom H (2014) Audio-visual classification and detection of human manipulation actions. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. DOI:10.1109/iros.2014.6942983.

Pirsiavash H and Ramanan D (2012) Detecting activities of daily living in first-person camera views. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, pp. 2847–2854. URL https://archive.ics.uci.edu/ml/datasets/.

Pitti A, Alirezaei H and Kuniyoshi Y (2009) Cross-modal and scale-free action representations through enaction. *Neural Networks* 22(2): 144–154. DOI:10.1016/j.neunet.2009.01.007.

Rahman SA, Song I, Leung M, Lee I and Lee K (2014) Fast action recognition using negative space features. *Expert Systems with Applications* 41(2): 574–587. DOI:10.1016/j.eswa.2013.07.082.

Rahman SA, Song I and Leung MKH (2012) Negative Space Template: A Novel Feature to Describe Activities in Video. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. pp. 1–7.

Ramirez-Amaro K, Beetz M and Cheng G (2017) Transferring skills to humanoid robots by extracting semantic representations from observations of human activities. *Artificial Intelligence* 247: 95–118. DOI:10.1016/j.artint.2015.08.009.

Razzaghi P, Palhang M and Gheissari N (2012) A new invariant descriptor for action recognition based on spherical harmonics. *Pattern Analysis and Applications* 16(4): 507–518. DOI:10.1007/s10044-012-0274-x.

Regneri M, Rohrbach M, Wetzel D, Thater S, Schiele B and Pinkal M (2013) Grounding action descriptions in videos. *Transactions of the Association of Computational Linguistics* 1: 25–36.

Reng L, Moeslund TB and Granum E (2005) Finding motion primitives in human body gestures. In: *International Gesture Workshop*. Springer, pp. 133–144.

Rizzolatti G and Craighero L (2004) The Mirror-neuron System. *Annu. Rev. Neurosci.* 27: 169–192.

Rizzolatti G and Luppino G (2001) The cortical motor system. *Neuron* 31(6): 889–901.

Rizzolatti G and Matelli M (2003) Two different streams form the dorsal visual system: anatomy and functions. *Experimental Brain Research* 153(2): 146–157. DOI:10.1007/s00221-003-1588-0. URL https://doi.org/10.1007/s00221-003-1588-0.

Rodriguez MD, Ahmed J and Shah M (2008) Action mach a spatio-temporal maximum average correlation height filter for action recognition. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, pp. 1–8. URL http://www.cs.ucf.edu/vision/public_html/data.html.

Roh MC, Shin HK and Lee SW (2010) View-independent human action recognition with volume motion template on single stereo camera. *Pattern Recognition Letters* 31(7): 639–647. DOI:10.1016/j.patrec.2009.11.017.

Roland PE, Larsen B, Lassen NA and Skinhoj E (1980a) Supplementary motor area and other cortical areas in organization of voluntary movements in man. *Journal of Neurophysiology* 43(1): 118–136. DOI:10.1152/jn.1980.43.1.118.

Roland PE, Skinhoj E, Lassen NA and Larsen B (1980b) Different cortical areas in man in organization of voluntary movements in extrapersonal space. *Journal of Neurophysiology* 43(1): 137–150. DOI:10.1152/jn.1980.43.1.137.

Rosales R and Sclaroff S (2003) A framework for heading-guided recognition of human activity. *Computer Vision and Image Understanding* 91(3): 335–367. DOI:10.1016/s1077-3142(03)00096-1.

Rosenbaum DA, Meulenbroek RG and Vaughan J (2001) Planning reaching and grasping movements: theoretical premises and practical implications. *Motor control* 5(2): 99–115.

Rosenbaum DA, Vaughan J, Barnes HJ and Jorgensen MJ (1992) Time course of movement planning: selection of handgrips for object manipulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(5): 1058.

Rosenfeld A and Ullman S (2016) Hand-object interaction and precise localization in transitive action recognition. In: *2016 13th Conference on Computer and Robot Vision (CRV)*. IEEE. DOI:10.1109/crv.2016.27.

Roshtkhari MJ and Levine MD (2012) A multi-scale hierarchical codebook method for human action recognition in videos using a single example. In: *2012 Ninth Conference on Computer and Robot Vision*. IEEE. DOI:10.1109/crv.2012.32.

Roshtkhari MJ and Levine MD (2013) Human activity recognition in videos using a single example. *Image and Vision Computing* 31(11): 864–876. DOI:10.1016/j.imavis.2013.08.005.

Rudolph M, Muhlig M, Gienger M and Bohme HJ (2010) Learning the consequences of actions: Representing effects as feature changes. In: *2010 International Conference on Emerging Security Technologies*. IEEE. DOI:10.1109/est.2010.9.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC and Fei-Fei L (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* 115(3): 211–252.

Russell SJ and Norvig P (2016) *Artificial Intelligence: A Modern Approach*. 3 edition. Pearson Education.

Ryan RM and Deci EL (2000) Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology* 25(1): 54–67.

Ryan S and Andreae J (1993) Learning sequential and continuous control. In: *Proceedings 1993 The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*. IEEE Comput. Soc. Press. DOI:10.1109/annes.1993.323019.

Ryoo MS and Aggarwal JK (2010) UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA). http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html.

Salih AAA and Youssef C (2016) Spatiotemporal representation of 3d skeleton joints-based action recognition using modified spherical harmonics. *Pattern Recognition Letters* 83: 32–41. DOI:10.1016/j.patrec.2016.05.032.

Sanchez-Riera J, Čech J and Horaud R (2012) Action recognition robust to background clutter by using stereo vision. In: *Computer Vision – ECCV 2012. Workshops and Demonstrations*. Springer Berlin Heidelberg, pp. 332–341. DOI:10.1007/978-3-642-33863-2_33.

Sanmohan and Krüger V (2009) Primitive based action representation and recognition. In: *Image Analysis*. Springer Berlin Heidelberg, pp. 31–40. DOI:10.1007/978-3-642-02230-2_4.

Sapienza M, Cuzzolin F and Torr PH (2013) Learning discriminative space–time action parts from weakly labelled videos. *International Journal of Computer Vision* 110(1): 30–47. DOI:10.1007/s11263-013-0662-8.

Schenatti M, Natale G Lorenzo andMetta and Sandini G (2003) Object grasping data-set. URL http://www.lira.dist.unige.it/. University of Genova, Italy.

Schüldt C, Laptev I and Caputo B (2004) Recognizing human actions: a local svm approach. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3. IEEE, pp. 32–36. URL http://www.nada.kth.se/cvap/actions/.

Science I and Lab S (2017) Sysu 3d human-object interaction set (sysu 3dhoi). https://1drv.ms/u/s!AryuLmtQeBD1gQ_RKG90_sGw7UK1.

Searle JR, Dennett DC and Chalmers DJ (1997) *The Mystery of Consciousness*. New York Review of Books.

Seidenari L, Varano V, Berretti S, Bimbo A and Pala P (2013) Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. pp. 479–485.

Shah D, Falco P, Saveriano M and Lee D (2016) Encoding human actions with a frequency domain approach. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. DOI:10.1109/iros.2016.7759780.

Shan Y, Zhang Z and Huang K (2015) Learning skeleton stream patterns with slow feature analysis for action recognition. In: *Computer Vision - ECCV 2014 Workshops*. Springer International Publishing, pp. 111–121. DOI:10.1007/

978-3-319-16199-0_8.

She L, Cheng Y, Chai JY, Jia Y, Yang S and Xi N (2014) Teaching robots new actions through natural language instructions. In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE. DOI:10.1109/roman.2014.6926362.

Shimozaki M and Kuniyoshi Y (2003) Integration of spatial and temporal contexts for action recognition by self organizing neural networks. In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*. IEEE. DOI:10.1109/iros.2003.1249227.

Silva C and Ribeiro B (2003) Navigating mobile robots with a modular neural architecture. *Neural Computing & Applications* 12(3-4): 200–211. DOI:10.1007/s00521-003-0383-y.

Singh S, Velastin SA and Ragheb H (2010) Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In: *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*. IEEE, pp. 48–55. URL http://velastin.dynu.com/MuHAVi-MAS/.

Sokolov AA, Gharabaghi A, Tatagiba MS and Pavlova M (2010) Cerebellar engagement in an action observation network. *Cerebral Cortex* 20(2): 486–491. DOI:10.1093/cercor/bhp117. URL http://dx.doi.org/10.1093/cercor/bhp117.

Soomro K, Zamir AR and Shah M (2012) UCF101: A dataset of 101 human actions classes from videos in the wild. *CoRR* abs/1212.0402. URL http://arxiv.org/abs/1212.0402.

Sorokin I, Seleznev A, Pavlov M, Fedorov A and Ignateva A (2015) Deep Attention Recurrent Q-Network. In: *NIPS Workshop on Deep Reinforcement Learning*. Montreal, Canada: Curran Associates, Inc.

Steels L (2003) Intelligence with Representation. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 361(1811): 2381–2395.

Stein S and McKenna SJ (2013) Combining embedded accelerometers with computer vision for recognizing food preparation activities. In: *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, pp. 729–738.

Su YC and Grauman K (2016) Leaving some stones unturned: Dynamic feature prioritization for activity detection in streaming video. In: *Computer Vision – ECCV 2016*. Springer International Publishing, pp. 783–800. DOI:10.1007/978-3-319-46478-7_48.

Sugita Y and Tani J (2008) A sub-symbolic process underlying the usage-based acquisition of a compositional representation: Results of robotic learning experiments of goal-directed actions. In: *2008 7th IEEE International Conference on Development and Learning*. IEEE. DOI:10.1109/devlrn.2008.4640817.

Sun Q and Liu H (2013) Action disambiguation analysis using normalized google-like distance correlogram. In: *Computer Vision – ACCV 2012*. Springer Berlin Heidelberg, pp. 425–437. DOI:10.1007/978-3-642-37431-9_33.

Sun Q, Liu H, Ma L and Zhang T (2016) A novel hierarchical bag-of-words model for compact action representation. *Neurocomputing* 174: 722–732. DOI:10.1016/j.neucom.2015.

09.074.

Sung J, Ponce C, Selman B and Saxena A (2012) Unstructured human activity detection from rgbd images. In: *Robotics and Automation (ICRA), 2012 IEEE International Conference on.* IEEE, pp. 842–849.

Tenorth M and Beetz M (2012) A unified representation for reasoning about robot actions, processes, and their effects on objects. In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems.* IEEE. DOI:10.1109/iros.2012. 6385529.

Tessitore G, Prevete R, Catanzariti E and Tamburrini G (2010) From motor to sensory processing in mirror neuron computational modelling. *Biological Cybernetics* 103(6): 471–485. DOI: 10.1007/s00422-010-0415-5.

Theodoridis T, Agapitos A, Hu H and Lucas SM (2008) Ubiquitous robotics in physical human action recognition: A comparison between dynamic ANNs and GP. In: *2008 IEEE International Conference on Robotics and Automation.* IEEE. DOI:10.1109/ robot.2008.4543676.

Thurau C (2007) Behavior histograms for action recognition and human detection. In: *Human Motion – Understanding, Modeling, Capture and Animation.* Springer Berlin Heidelberg, pp. 299–312. DOI:10.1007/978-3-540-75703-0_21.

Thurau C and Hlaváč V (2007) n-grams of action primitives for recognizing human behavior. In: *Computer Analysis of Images and Patterns.* Springer Berlin Heidelberg, pp. 93–100. DOI: 10.1007/978-3-540-74272-2_12.

Thurau C and Hlaváč V (2009) Recognizing human actions by their pose. In: *Lecture Notes in Computer Science.* Springer Berlin Heidelberg, pp. 169–192. DOI:10.1007/978-3-642-03061-1_ 9.

Tucker M and Ellis R (1998) On the Relations between seen Objects and Components of Potential Actions. *Journal of Experimental Psychology: Human Perception and Performance* 24(3): 830–846.

Tunik E, Frey S and T Grafton S (2005) Virtual lesions of the anterior intraparietal area disrupt goal-dependent on-line adjustments of grasp 8: 505–11.

Vafeias E and Ramamoorthy S (2014) Joint classification of actions and object state changes with a latent variable discriminative model. In: *2014 IEEE International Conference on Robotics and Automation (ICRA).* IEEE. DOI:10.1109/icra.2014. 6907570.

Vanderelst D and Winfield A (2017) Rational imitation for robots: the cost difference model. *Adaptive Behavior* 25(2): 60–71. DOI:10.1177/1059712317702950.

Veeraraghavan A, Chellappa R and Roy-Chowdhury AK (2006) The function space of an activity. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on,* volume 1. IEEE, pp. 959–968.

Vieira AW, Nascimento ER, Oliveira GL, Liu Z and Campos MF (2014) On the improvement of human action recognition from depth map sequences using space–time occupancy patterns. *Pattern Recognition Letters* 36: 221–227. DOI:10.1016/j. patrec.2013.07.011.

Vieira AW, Nascimento ER, Oliveira GL, Liu Z and Campos MFM (2012) STOP: Space-time occupancy patterns for 3d action recognition from depth map sequences. In: *Progress in Pattern Recognition, Image Analysis, Computer Vision, and*

Applications. Springer Berlin Heidelberg, pp. 252–259. DOI: 10.1007/978-3-642-33275-3_31.

Viet VH, Ngoc LQ, Son TT and Hoang PM (2015) Multiple kernel learning and optical flow for action recognition in RGB-d video. In: *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE).* IEEE. DOI: 10.1109/kse.2015.39.

Vitkute-Adzgauskiene D, Markievicz I, Krilavicius T, Tamosiunaite M, Kulvicius T, Bodenhagen L and Langer H (2014) Chemlab corpus. EU-FP7-STREP (600578) ACAT, Learning and Execution of Action Categories, D1.1: Text corpora and image databases.

von Goethe JW (1808) *Faust, Eine Tragödie.* Cotta, Tübingen.

Vosgerau G (2009) *Mental Representation and Self-Consciousness: From Basic Self-Representation to Self-Related Cognition.* Mentis.

Vuga R, Aksoy EE, Wörgötter F and Ude A (2015) Probabilistic semantic models for manipulation action representation and extraction. *Robotics and Autonomous Systems* 65: 40–56. DOI: 10.1016/j.robot.2014.11.012.

Wang J, Liu Z, Wu Y and Yuan J (2012) Mining action-let ensemble for action recognition with depth cameras. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.* IEEE, pp. 1290–1297. URL http://research.microsoft.com/ en-us/um/people/zliu/actionrecorsrc/.

Wang L, Li R and Fang Y (2017) Power difference template for action recognition. *Machine Vision and Applications* 28(5-6): 463–473. DOI:10.1007/s00138-017-0848-0.

Webb TW, Kean HH and Graziano MSA (2016) Effects of Awareness on the Control of Attention. *Journal of Cognitive Neuroscience* 28(6): 842–851.

Wechsler H, Duric Z and Li F (2002) Hierarchical interpretation of human activities using competitive learning. In: *Object recognition supported by user interaction for service robots.* IEEE Comput. Soc. DOI:10.1109/icpr.2002.1048308.

Weiller D (2010) Unsupervised learning of reflexive and action-based affordances to model adaptive navigational behavior. *Frontiers in Neurorobotics* 4. DOI:10.3389/fnbot.2010.00002.

Weinland D, Ronfard R and Boyer E (2006) Free viewpoint action recognition using motion history volumes. *Computer vision and image understanding* 104(2-3): 249–257. URL http:// 4drepository.inrialpes.fr/public/datasets.

Weinland D, Ronfard R and Boyer E (2011) A survey of vision-based methods for action representation, segmentation and recognition. *Computer Vision and Image Understanding* 115(2): 224–241.

Wermter S, Weber C, Elshaw M, Gallese V and Pulvermüller F (2005) Grounding neural robot language in action. In: *Biomimetic Neural Learning for Intelligent Robots.* Springer Berlin Heidelberg, pp. 162–181. DOI:10.1007/11521082_10.

Whiten C, Laganiere R and Bilodeau GA (2013) Efficient action recognition with MoFREAK. In: *2013 International Conference on Computer and Robot Vision.* IEEE. DOI: 10.1109/crv.2013.30.

Wolpert DM and Kawato M (1998) Multiple Paired Forward and Inverse Models for Motor Control. *Neural networks* 11(7-8): 1317–1329.

Worgotter F, Aksoy EE, Kruger N, Piater J, Ude A and Tamosiunaite M (2013) A simple ontology of manipulation

actions based on hand-object relations. *IEEE Transactions on Autonomous Mental Development* 5(2): 117–134. DOI: 10.1109/tamd.2012.2232291.

Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X and Xiao J (2015) 3D ShapeNets: A Deep Representation for Volumetric Shapes. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1912–1920.

Xi N and Tarn TJ (1999) *Journal of Intelligent and Robotic Systems* 25(4): 281–293. DOI:10.1023/a:1008183129849.

Xia L, Chen CC and Aggarwal JK (2012) View invariant human action recognition using histograms of 3d joints. In: *Computer vision and pattern recognition workshops (CVPRW), 2012 IEEE computer society conference on*. IEEE, pp. 20–27.

Xiao Q and Cheng J (2013) Human action recognition framework by fusing multiple features. In: *2013 IEEE International Conference on Information and Automation (ICIA)*. IEEE. DOI:10.1109/icinfa.2013.6720438.

Yang J, Xu Y and Chen C (1997) Human action learning via hidden markov model. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 27(1): 34–44. DOI: 10.1109/3468.553220.

Yang Y, Guha A, Fermuller C and Aloimonos Y (2014) Manipulation action tree bank: A knowledge resource for humanoids. In: *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE. DOI:10.1109/humanoids.2014.7041483.

Yao B, Jiang X, Khosla A, Lin AL, Guibas L and Fei-Fei L (2011) Human action recognition by learning bases of action attributes and parts. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, pp. 1331–1338. URL http://vision.stanford.edu/Datasets/40actions.html.

Zambelli M and Demiris Y (2017) Online multimodal ensemble learning using self-learned sensorimotor representations. *IEEE Transactions on Cognitive and Developmental Systems* 9(2): 113–126. DOI:10.1109/tcds.2016.2624705.

Zampogiannis K, Yang Y, Fermuller C and Aloimonos Y (2015) Learning the spatial semantics of manipulation actions through preposition grounding. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. DOI:10.1109/icra.2015.7139371.

Zech P, Haller S, Rezapour Lakani S, Ridge B, Ugur E and Piater J (2017) Computational Models of Affordance in Robotics: A Taxonomy and Systematic Classification. *Adaptive Behavior* 25(5): 235–271.

Zhang C, Zhang H, Guo R and Parker LE (2016) Unified robot learning of action labels and motion trajectories from 3d human skeletal data. In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE. DOI:10.1109/roman.2016.7745095.

Zhang J and Zhuang Y (2007) View-independent human action recognition by action hypersphere in nonlinear subspace. In: *Advances in Multimedia Information Processing ˘ PCM 2007*. Springer Berlin Heidelberg, pp. 108–117. DOI:10.1007/978-3-540-77255-2_13.

Zhang Z, Hu Y, Chan S and Chia LT (2008) Motion context: A new representation for human action recognition. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 817–829. DOI:10.1007/978-3-540-88693-8_60.

Zhou Z, Song M, Zhang L, Tao D, Bu J and Chen C (2011) kPose: A new representation for action recognition. In: *Computer Vision – ACCV 2010*. Springer Berlin Heidelberg, pp. 436–447. DOI: 10.1007/978-3-642-19318-7_34.

Zhu G, Zhang L, Shen P and Song J (2016) Human action recognition using multi-layer codebooks of key poses and atomic motions. *Signal Processing: Image Communication* 42: 19–30. DOI:10.1016/j.image.2016.01.003.

Zimmer M and Doncieux S (2018) Bootstrapping q -learning for robotics from neuro-evolution results. *IEEE Transactions on Cognitive and Developmental Systems* 10(1): 102–119. DOI: 10.1109/tcds.2016.2628817.

Author Biographies

Philipp Zech is a postdoctoral researcher at the University of Innsbruck. He received his Ph.D. degree in Computer Science from the University of Innsbruck in 2014. He is interested in developmental and cognitive robotics, affordance learning and intelligent and adaptive manipulation. He has already published more than 30 papers in international journals and conferences, two of which have received best-paper awards. He recently joined the editorial board of Adaptive Behavior as an associate editor.

Erwan Renaudo is a postdoctoral researcher at the University of Innsbruck. He completed his PhD degree in 2016 at Pierre and Marie Curie University where he worked on bio-inspired robotic architectures with ensemble reinforcement learning. His research interests focus on learning methods for autonomous robots. He is particularly interested in autonomous behavior generation, from action learning to coordination of habitual and goal-directed behaviors.

Simon Haller is with the Department of Computer Science, University of Innsbruck, from where he received his B.Sc. in 2006. In 2012 he was granted the professional title Ing. by the Federal Ministry of Economics, Austria. He is a researcher and a scientific systems engineer. His work focuses on problem solving in the field of robotic hard- and software.

Xiang Zhang Xiang Zhang is currently a PhD student at University of Innsbruck. He received his M.Sc. degree in System and Control from Delft University of Technology in 2016. He is interested in robot learning from demonstrations and intelligent and adaptive manipulation.

Justus Piater Piater is a professor of computer science at the University of Innsbruck, Austria, where he leads the Intelligent and Interactive Systems group. He holds a M.Sc. degree from the University of Magdeburg, Germany, and M.Sc. and Ph.D. degrees from the University of Massachusetts Amherst, USA, all in computer science. Before joining the University of Innsbruck in 2010, he was a visiting researcher at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, a professor of computer science at the University of Liège, Belgium, and a Marie-Curie research fellow at GRAVIR-IMAG, INRIA Rhône-Alpes, France. His research interests focus on visual perception, learning and inference in sensorimotor systems. He has published more than 170 papers in international journals and conferences, several of which have received best-paper awards, and currently serves as Associate Editor of the IEEE Transactions on Robotics.