

# Traversability: A Case Study for Learning and Perceiving Affordances in Robots

Emre Ugur<sup>1,2,3</sup> and Erol Şahin<sup>1</sup>

<sup>1</sup> METU, Dept. of Computer Engineering, KOVAN Res. Lab., Ankara, Turkey

<sup>2</sup> NICT, Biological ICT Group, Kyoto, Japan

<sup>3</sup> ATR, Computational Neuroscience Labs., Kyoto, Japan

April 28, 2010

**Running title:** Traversability

**Contact person:** Asst.Prof.Dr. Erol Şahin

Dept. of Computer Engineering

Middle East Technical University

İnönü Bulvarı, Ankara, 06531, Turkey

E-mail: erol@ ceng.metu.edu.tr

Phone: +90-312-210 5539

Fax: +90-312-210 5544

## Abstract

The concept of affordances, introduced by J.J. Gibson in Psychology, has recently attracted interest in autonomous robotics towards the development of cognitive systems. In earlier work (Sahin et al., Adaptive Behavior, vol.15(4), pp. 447-472, 2007), we reviewed the uses of this concept in different fields and proposed a formalism to use affordances at different levels of robot control. In this paper, we first review studies in Ecological Psychology on the learning and perception of traversability in organisms and describe how the existence of traversability was judged to exist. Then, we describe the implementation of one part of the affordance formalism for the learning and perception of traversability affordances on a mobile robot equipped with range sensing ability. Through experiments inspired from Ecological Psychology, we show that the robot, by interacting with its environment, can learn to perceive the traversability affordances. Moreover, we claim that three of the main attributes that are commonly associated with affordances; that is, affordances being relative to the environment, providing perceptual economy and providing general information, are simply consequences of learning from the interactions of the robot with the environment.

**Keywords:** Affordance, traversability, range image, perception, learning, autonomous robots,

# 1 Introduction

The concept of affordances was conceived by the famous psychologist J.J. Gibson during his studies in Psychology while he was developing a “theory of information pick-up” by organisms:

“The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment.” (J. J. Gibson, 1979/1986, p. 127)

The concept, described through inspirational but also vague discussions by J.J. Gibson such as the one quoted above, turned out to be very influential and has attracted interest from a wide range of fields, ranging from Neuroscience and Human Computer Interaction to Autonomous Robotics. In our earlier work (Şahin, Çakmak, Doğar, Uğur, & Üçoluk, 2007), we summarized the context behind the conception of the concept, speculated on its evolution within J.J. Gibson’s studies, and reviewed the usage of affordances in different fields. We concluded that the confusion surrounding the concept had stemmed from that fact that J.J. Gibson’s own ideas on the concept were not finalized during his lifetime and was left in an ambiguous state.

A direct consequence of this confusion was the association of different attributes to the concept of affordances in different contexts. Some claimed that affordances are relations within the organism-environment system, whereas others used affordances to refer to the perceptual clues that determine its usage, or even to the actions that can be associated with a certain object. Although each claim supported their perspective using quotations from J.J. Gibson’s own writings, the set of attributes that were associated to affordances painted a patchy, and sometimes contradictory view.

In this paper, we will consider three of the main attributes<sup>1</sup> that are commonly associated with affordances in robotics; namely,

- *Affordances are relative.* This argument, generally accepted within most contexts, is usually linked to the complementarity of the organism and the environment. According to this view, the existence of an affordance is neither defined by the environment nor by the organism alone but through their interaction. For instance, the climb-ability of a stair step is not only determined by the metric measure of the step height, but also by one’s leg-length.
- *Affordances provide perceptual economy.* The concept of affordances is often used as support for minimality in perception to argue that one do not have to perceive all the qualities of their environment in order to accomplish a simple task such as wandering around. In this sense, one would directly perceive the traversability of a path without recognizing the objects on its path and making additional “mental inferences” over them.
- *Affordances provide general information.* The discussion on affordances are mostly based on the general relations that pertain to the interaction of the organism with its environment such as sit-ability, climb-ability, and cross-ability. It is usually assumed that the use of affordances enables one to deduce whether a designer’s chair that he sees for the first time would support sittability, or whether a coconut shell can be used to carry water in the place of a cup.

Specifically, we use the formalism that extended the Gibsonian view of the concept to use at different levels of autonomous robot control (Şahin et al., 2007) as a framework, and propose a method for learning affordances on robots to show that how these attributes can be obtained.

## 2 Traversability

The studies in Ecological Psychology focused on the perception, learning and use of affordances in organisms, whereas the studies in Autonomous Robotics aimed to discover how this concept could

---

<sup>1</sup>Although these arguments are certainly inspired from J.J. Gibson’s own writings, we refrain from attributing them to him, in order to avoid the discussion of what he actually meant (or did not mean) in his writings.

be utilized in controlling robots. However, despite working on the same concept, most of these studies have preferred to cite J.J. Gibson’s own writings on affordances, neglecting each other.

In this paper, we link the studies within these two fields through a case study on traversability. The verb “traverse” is defined as “to pass or move over, along, or through”. Hence traversability refers to the affordance of being able to traverse. The learning and perception of traversability is a fundamental competence for both organisms and autonomous mobile robots since most of their actions depend on their mobility.

In Ecological Psychology, the learning and perception of traversability in organisms is probably one of the well-studied topics on affordances. Although it is not known precisely which visual cues are actually used in space perception (Sedgwick, 2001), many organisms are known to use visual perception to detect whether the environment’s spatial layout allows them to carry out their locomotor activities, such as crawling, walking or jumping.

Although traversability is considered a fundamental capability for mobile robots, it has long been limited to the problem of simple obstacle avoidance where the robot tries to avoid making any physical contact with the environment, and heads only to open spaces. In general, proximity sensors are employed to detect whether there is an object or not. When such approaches are used, the robot’s response would be the same whether it encounters an unpenetrable wall or a balloon that can just be pushed aside without any damage. A stair that is traversable for a hexapod robot may not be traversable for a wheeled one. Similarly a white vertical flat surface may be an impenetrable wall in one environment whereas in another environment a similar surface may be a door that can just be pushed to open. Therefore, a method that can automatically learn the traversability affordances from the robot’s interactions with the environment would be valuable for robotics.

## 2.1 Ecological Psychology

Simple amphibians are known to perceive whether varying size barriers, apertures, overhead holes and pits afford locomotion or not. Toads, known to possess depth perception through stereopsis (Collett, 1977) tend to walk into shallow pits, and jump over deeper ones (Lock & Collett, 1979). The relation between the aperture and their body width is complex, since the dynamics of the interaction also depend on the orientation of the gap and animal’s jumping direction. But leopard frogs for example, when challenged with a stimulus at their rear, are shown to jump through apertures that are larger than their own bodies (Ingle & Cook, 1977).

Infants with crawling or walking ability are able to detect more complex affordances such as traversability of rigid/non-rigid surfaces and act accordingly (E. J. Gibson et al., 1987). They can use both haptic and visual information provided by the environment and perceive the traversability affordance implicitly taking into account their mode of locomotion. There also exist situations where haptic and visual cues are contradictory. For example in the so-called “visual cliff” experiments (E. Gibson & R.D.Walk, 1960), crawling infants are placed on a glass surface part of which is placed on a table covered with a textured sheet, whereas the remaining part is kept on air (supported from only its sides). Thus although the glass is a rigid surface, the part on the table gives appearance of solidity, and the other part becomes a visual cliff. In such situations, crawling infants tend not to go over the apparently unsupported surface even if their mothers call them from the other side.

In experiments conducted with human adults, subjects were queried on the existence (or non-existence) of affordances. Warren’s stair-climbing experiments (Warren, 1984), have generally been accepted as a seminal work on the analysis of affordances. According to Warren, “to determine whether a given path affords locomotion, the behaviorally relevant properties of the environment must be analyzed in relation to relevant properties of the animal and its action system”. Thus a specific set of values of the functionally related variables is identified for the activity of stair-climbing. Since the environment should be perceived in terms of *intrinsic* or *body-scaled* metrics, not in absolute or global dimensions, this specific set of values is expressed as  $\pi$ , a dimensionless ratio of animal property (leg-length) and environment property (stair-height). The particular value of these ratios that signaled the existence of an affordance were called the *critical points*. It was argued that critical points remain constant across humans with different body sizes and provide a natural basis for perceptual categories (e.g. categories of climb-able and not-climb-able).

Warren’s studies were followed by studies that further explore the underlying mechanisms of traversability in different environments and that identify the visual channels and cues in human affordance perception. For example slanted surfaces are included into the environment in (Kinsella-Shaw, Shaw, & Turvey, 1992) and human subjects are shown to correctly predict the walk-on-ability affordance of slopes when they perceive them at a distance. The roles of optical and geographical slants which imply relative and absolute measures are discussed in the detection of these affordances. In another work, (Warren & Whang, 1987) studied which properties of the environment and human body are used in visual guidance for walking through apertures. The constant  $\pi$  ratio is defined as a proportion of the environment-related variable *aperature width* and action-related organism variable *shoulder width*. In the experiments, where subjects are asked to judge whether they can pass through apertures without rotating their shoulders, the predicted critical  $\pi$  ratio is found to be compatible with the real one, that is found by actually executing the actions. It was further shown that static human subjects with monocular vision looking through a reduction screen are as successful as moving subjects with binocular vision in the detection of pass-through-able apertures. Thus, stereo vision and optic flow are not necessarily involved in the process of traversability perception, however “perceived eyeheight” as an intrinsic measure is shown to be used. Traversability is also studied in environments with barriers (Marcilly & Luyat, 2008), where human subjects are asked to judge the pass-under-ability of barriers at different heights. The predictions of the subjects are found to be valid as in previous experiments and compatible with the constant critical ratio  $\pi$ , defined as the proportion of subject-to-barrier height. Instead of passing-under, when the subjects are asked whether they can walk-over obstacles (Cornus, Montagne, & Laurent, 1999) or gaps (Chemero, Klein, & Cordeiro, 2003) of different sizes, the traversability detection is found to be successful as well.

In summary the studies discussed above are generally used to show that organisms can perceive whether the surface layout affords traversability or not and identified a dimensionless ratio between the properties of the organism and environment related to the action as a proof of existence.

### 2.1.1 Learning of affordances

J.J. Gibson was not particularly interested in development and “his concern was with perception” (Szokolszky, 2003) only. As a result, he did not discuss the concept of affordances from a developmental point of view, and only mentioned that affordances are learned in childhood (J. J. Gibson, 1986). It is generally accepted that infants’ exploration, through physical interaction with the environment, is very important in development of locomotion related perceptual and motor skills (Adolph, Bertenthal, Boker, Goldfield, & Gibson, 1997). E.J. Gibson argued that learning is neither the construction of representations from smaller pieces, nor the association of a response to a stimulus. Instead, she claimed, learning is “discovering *distinctive* features and *invariant* properties of things and events” (E. J. Gibson, 2000) or “discovering the information that specifies an affordance” (E. J. Gibson, 2003). Learning is not “enriching the input”, but discovering the critical perceptual information in that input. She named this process of discovery *differentiation*, and defined it as a “narrowing down from a vast manifold of (perceptual) information to the minimal, optimal information that specifies the affordance of an event, object, or layout” (E. J. Gibson, 2003). The method proposed for the learning and perception of affordances is mainly inspired by E.J. Gibson’s studies.

## 3 Affordance formalization

In a previous work (Şahin et al., 2007), we proposed a new formalization of affordances, based partially on Chemero’s formalization (Chemero, 2003) and outlined how affordances can be used at different levels of robot control, ranging from perception and learning to planning. In our formalization we argue that an affordance is an acquired relation that can be represented as a nested triplet

$$(effect, (entity, behavior))$$

indicating that when the agent applies the *behavior* on the *entity* the *effect* is generated. Here, the term *entity* denotes the environmental relata of the affordance and represents the initial state of



Figure 1: **Left:** The KURT3D mobile robot platform. **Center:** KURT3D simulated in MACSim. **Right:** A 3D range image obtained from the KURT3D.

the environment (including the perceptual state of the agent) as perceived by the agent. Although for some affordances the term object would perfectly encapsulate the environmental relata, for others, the relata may be too complex to be confined to an object - such as the layout among multiple objects. In the rest of the paper, we will freely use object instead of *entity* for the sake of clarity. The term *behavior* denotes the agent’s relata which represents the part of the agent that is generating the interaction with the environment that actualized the affordance. It consists of the agent’s embodiment that generates the perception-action loop that can realize the affordance. Finally, the term *effect* denotes the effect that is generated by the agent’s execution of the *behavior* on the *entity*. More specifically, a certain behavior applied to a certain entity should produce a certain *effect*. For instance, the lift-ability affordance implicitly assumes that, when the *lift* behavior is applied to a *can*, it produces the effect *lifted*, meaning that the can’s position, as perceived by the agent, is elevated.

Based on these arguments, we argue that through its interactions with a can, a robot can acquire *relation instances* of the form:

$$(lifted, (black-can, lift-with-right-hand))$$

meaning that there exists a potential to generate the effect *lifted* when *lift-with-right-hand* is applied to *black-can*. Note that the terms *black-can*, *lifted* and *lift-with-right-hand* are used just as short-hand labels to denote the perceptual representations of the environment and the agent. For instance the representation of black can be a raw feature vector derived from all the sensors of the robot looking at the *black-can* before it attempts to apply its lift behavior.

Arguing that affordances should be relations with predictive abilities, rather than a set of unconnected relation instances, we proposed a process of generating *equivalence classes* that can be applied on this representation. For instance, a robot can achieve the effect *lifted*, by applying the *lift-with-right-hand* behavior on a *black-can*, or a *blue-can*. It can thus learn a relation:

$$(lifted, (<*-can>, lift-with-right-hand))$$

where  $<*-can>$  denotes the derived invariants of the entity equivalence class.

The nesting inside the affordance triplet provided support for planning over learned affordances (Uğur, Oztop, & Şahin, 2009), and can be removed within the context of this study for simplicity as:

$$(lifted, <*-can>, lift-with-right-hand)$$

## 4 Experimental Framework

A mobile robot, equipped with a 3D laser range finder, and its physics-based simulator, is used as the experimental platform (Figure 1). The 3D laser scanner has a horizontal range of  $180^\circ$ , and is able to sweep a vertical range of  $\pm 82.8^\circ$ . The robot is simulated in MACSim (Uğur, Doğar, Soysal, Çakmak, & Şahin, 2006), a physics-based simulator built using the ODE (Open Dynamics Engine)<sup>2</sup> physics engine.

<sup>2</sup><http://ode.org/>

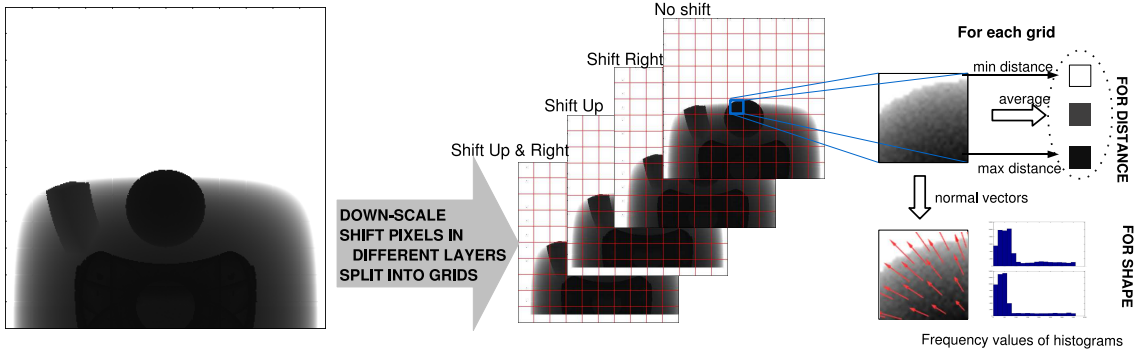


Figure 2: Perceptual processing of the range image.

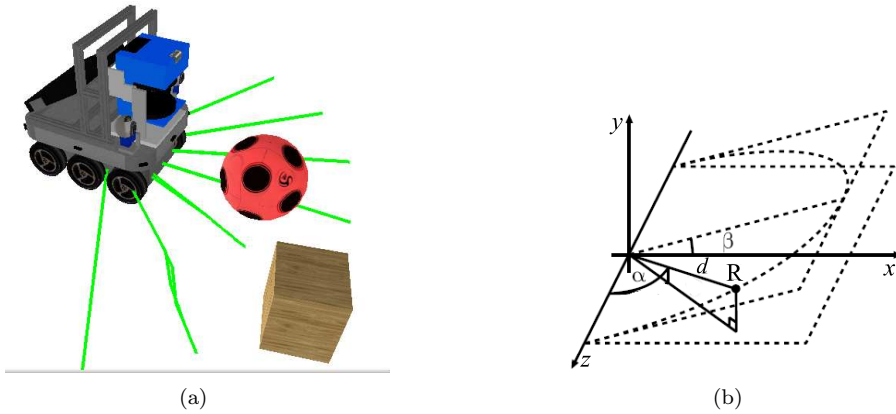


Figure 3: **Left:** The simulated robot and the trajectories recorded during the execution of its 7 different move behaviors. Note that in some cases, the robot’s movement is obstructed by the box. **Right:** The coordinate system of 3D scanning. The two planes correspond to the different 2-D slices sensed in 2D scanning. The laser beam reflected from R is transmitted at  $\alpha$  degrees in the scanning plane of the scanner which has a pitch angle of  $\beta$ .

## 4.1 Perception

The robot perceives the world through a range image of resolution  $720 \times 720$ . As sketched in Figure 2, the image is first down-scaled to  $360 \times 360$  pixels in order to reduce the noise, and split into uniform size squares in a grid. The grid squares are shifted in order to have a representation that provides overlaps. Finally, low-level generic features are extracted for each grid square where 3 features are related to distance characteristics of the grid square and 36 features to shape. The features of the different grid squares are then collected and stored in a large one-dimensional feature vector  $\mathbf{f}_k$  that represents the perception of the robot before  $k^{\text{th}}$  interaction.

The distance-related features of each grid square are defined as the minimum, maximum, and mean range values of that grid square. In order to derive the shape-related features, the position of each pixel in the range image (see Figure 3(b)) relative to the laser scanner is computed using:

$$\mathbf{p}_{r,c} = \begin{bmatrix} d_{r,c} \sin(\alpha_{r,c}) \cos(\beta_{r,c}) \\ d_{r,c} \sin(\alpha_{r,c}) \sin(\beta_{r,c}) \\ d_{r,c} \cos(\alpha_{r,c}) \end{bmatrix}$$

where  $d$  is the distance measured,  $r$  and  $c$  are the row and column indexes of the corresponding point, respectively. After finding the positions, the normal vector of the local surface around each point is computed by using the positions of the two neighbours in the range image:

$$\mathbf{N}_{r,c} = (\mathbf{p}_{r-n,c} - \mathbf{p}_{r,c}) \times (\mathbf{p}_{r,c-n} - \mathbf{p}_{r,c})$$

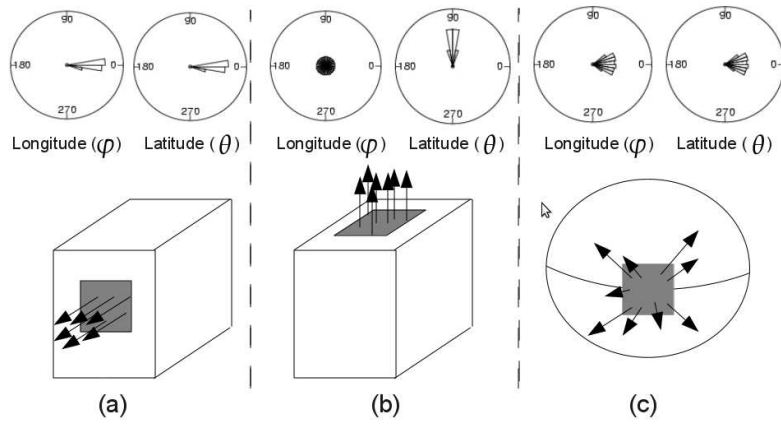


Figure 4: (a) Vertical, (b) horizontal and (c) spherical surface patches and their corresponding angular histograms of normal vectors in latitude ( $\theta$ ) and longitude ( $\varphi$ ). In (b), the orthogonal projection of the normal vectors onto the horizontal plane should create zero size vectors in ideal conditions and the angles in longitude in this situation should be undefined. However, the noise in sensing creates small random perturbations in these normal vectors which in turn results in randomly distributed angular histograms in longitude.

where  $n$  corresponds to the neighbour pixel distance and is set to 5. In spherical coordinates, the unit length 3D normal vector is represented by two angles, polar ( $\theta$ ) and azimuthal ( $\varphi$ ) angles that encode information along latitude and longitude, respectively. The polar angle ( $\theta$ ) corresponds to the angle between y-axis and the normal vector, whereas  $\varphi$  is the angle between z-axis and the normal vector's orthogonal projection on x-z plane. After polar and azimuthal angles are computed for each pixel, angular histograms are computed in both dimensions for each grid square and are sliced into 18 intervals of  $20^\circ$  each. At the end, frequency values of angular histograms are used as 36 shape related features. This representation encodes the distribution of the local surface normal vectors of the corresponding grid square as shown in Figure 4.

## 4.2 Interaction

The robot has seven pre-coded behaviors to move. The execution of a behavior,  $b^i$  where  $0 \leq i \leq 6$ , consists of first rotating the robot in place for a certain angle (one of  $0^\circ, \pm 20^\circ, \pm 40^\circ, \pm 60^\circ$ ), and then driving it forward for  $70cm$ , as shown in Figure 3(a). The robot measures its actual displacement and change in its orientation through its wheel encoders and uses the discrepancy between the predicted and measured values to detect whether it was able to traverse or not.

The interaction of the robot with its environment consists of episodes. In episode  $k$ , the robot first computes the feature vector  $\mathbf{f}_k$  for the environment to be acted upon. Then it executes behavior  $b^i$  and records the result ( $r_k^i$ ) as success or fail. During the *exploration phase*, which takes place in MACSim for obvious safety reasons, the robot executes all of its behaviors within a given environment and records its experience in the form of affordance relation instances as  $\langle r_k^i, \mathbf{f}_k, b^i \rangle$  triplets (Algorithm 1).

## 4.3 Learning

Learning consists of the formation of *entity equivalence classes* to perceive affordances. Specifically, for a given behavior, the robot discovers the invariant features of the environment for traversability (or non-traversability) and learns to map these invariant features to its affordances. The discovery of the invariant features and the learned mapping from these features to the affordances implements the *entity equivalence class* for traversability.

Learning is conducted as a batch process that takes place after the exploration. The learning phase consists of two steps as explained in Algorithm 2 and is carried out separately for each



---

**Algorithm 1** Exploration phase

---

- 1: **for** each trial  $k$  (from 1 to  $m$ ) **do**
  - 2:   Put the robot in a randomly constructed environment.
  - 3:   Make a 3D scan.
  - 4:   Compute feature vector,  $\mathbf{f}_k$ .
  - 5:   **for** each behavior  $b^i$  **do**
  - 6:     Perform  $b^i$ .
  - 7:     Find result of behavior,  $r_k^i$ .
  - 8:     Put  $\langle b^i, \mathbf{f}_k, r_k^i \rangle$  into repository.
  - 9:     Reset robot and object positions.
  - 10:   **end for**
  - 11: **end for**
- 

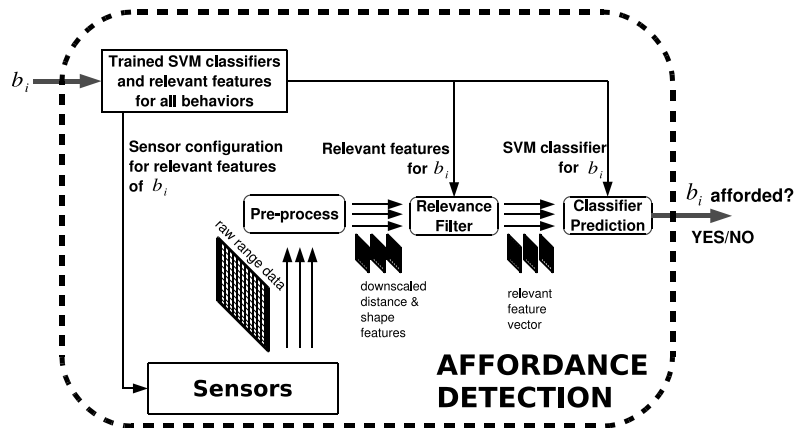


Figure 5: The affordance prediction module receives the behavior id (or direction) as input and predicts the behavior’s affordance based on the percept of the environment.

behavior. In the first step, the ReliefF method (Kira & Rendell, 1992; Kononenko, Simec, & Robnik-Sikonja, 1997) is used to automatically pick out the relevant features for the perception of traversability. Specifically, in ReliefF, the relevancy of a feature is increased, if the feature takes similar values for the situations that have same execution results, and it has different values for situations that have different results (see Algorithm 3). The features with relevancy values above a certain threshold are marked as relevant.

---

**Algorithm 2** Learning phase

---

- 1: **for** each behavior  $b^i$  **do**
  - 2:   Fetch samples  $\langle \mathbf{f}_k, r_k^i \rangle$  from repository for behavior  $b^i$ .
  - 3:   Find a set of relevant features  $\mathcal{F}^i$  using Algorithm 3.
  - 4:   Train the SVM model,  $M^i$ , with relevant features.
  - 5:   Store  $\mathcal{F}^i$  and  $M^i$  for perception of affordances in *execution mode*.
  - 6: **end for**
- 

During the second step, linear kernel SVMs (Support Vector Machines)<sup>3</sup>(Vapnik, 1998) are trained to classify traversability based on relevant features. These SVM classifiers are later used to predict the existence of affordances as illustrated in Figure 5.

<sup>3</sup>The LibSVM software that is used in this study, is available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

---

**Algorithm 3** Computation of feature weights for behavior  $b^i$ 

---

$n_f$ : number of features

$w_d$ : weight of  $d^{\text{th}}$  feature

$m$ : number of iterations, experimentally set to 1000

$f_l$ : the feature vector computed in  $l^{\text{th}}$  situation (interaction)





$f_l[d]$ : the normalized value of  $d^{\text{th}}$  feature computed in  $j^{\text{th}}$  situation (interaction)

```
1:  $w_d \leftarrow 0$ , where  $1 \leq d \leq n_f$ ,  $n_f$  is number of features (initialize weights)
2: for  $i = 0$  to  $m$  do
3:   Select a random feature vector  $f_l$  from  $\{ \langle r_k^i, f_k, b^i \rangle \}$ .
4:   Compute distance of  $f_l$  to all other samples in  $\{f_k\}$ .
5:   Find 10 feature vectors closest to  $f_l$  with execution results  $r_l^i$ , i.e. find the most similar situations to the  $l^{\text{th}}$  situation with the same result. Put them into set of nearest hits,  $\mathcal{H}$  ( $\mathcal{H} = \{f_{1'}, \dots, f_{10'}\}$ ).
6:   Find 10 nearest feature vectors with execution results different from  $r_l^i$ , and put them into set of nearest misses,  $\mathcal{M}$  ( $\mathcal{M} = \{f_{1''}, \dots, f_{10''}\}$ ).
7:   for  $d = 0$  to  $n_f$  do
8:      $w_d \leftarrow w_d - \frac{1}{m \cdot 10} \sum_{j=1}^{10} |f_l[d] - f_{j'}[d]| + \frac{1}{m \cdot 10} \sum_{j=1}^{10} |f_l[d] - f_{j''}[d]|$ 
9:   end for
10: end for
```

---

## 5 Experiments

The robot interacted in an environment filled with different types of objects listed below:

- rectangular boxes (  ) that are non-traversable.
- spherical objects (  ) that are traversable since they could roll in all directions,
- cylindrical objects in upright position (  ) that are non-traversable,
- cylindrical objects lying on the ground (  ), that may or may not be traversable,
- ramps, that may or may not be traversable,
- gaps, that may or may not be traversable.

Note that the description provided above is rather crude. First, the robot does not have the concept of an object, and our discussion at the level of objects is only to ease our discussion. Second, multiple objects can be present in the environment and the traversability is a complex function of not only the individual objects but also their layouts. Third, the size, relative placement and orientation of the objects vary during the experiments and hence their traversability.

### 5.1 Parameter optimization

Both the perceptual representation of the robot and the learning phase contains a number of parameters that needs to be optimized to obtain the best affordance prediction performance. Regarding the perceptual representation, the effect of grid size as well as the effect of overlapped (versus non-overlapped) grid representation needs to be decided. Regarding the learning phase, the relevancy threshold and the tolerance parameter to be used during the training of the SVM's needs to be optimized. In all the experiments reported below, unless otherwise stated, we carried out 5000 interactions during which the robot faced up to 12 objects (boxes, cylinders, and spheres of random sizes) that were placed at random locations and with different orientations. During the evaluation of prediction accuracies, the training set was split into 5, and 5-fold cross-validation was performed.

As a result of the optimization process (see Appendix), we used a representation with  $5 \times 5$  grid with 4 overlapping layers. Using these parameters, the feature vector  $f_k$  consists of  $4 \times (5 \times 5) \times 39 = 3900$  features. The parameters of the learning phase were optimized for each behavior separately.

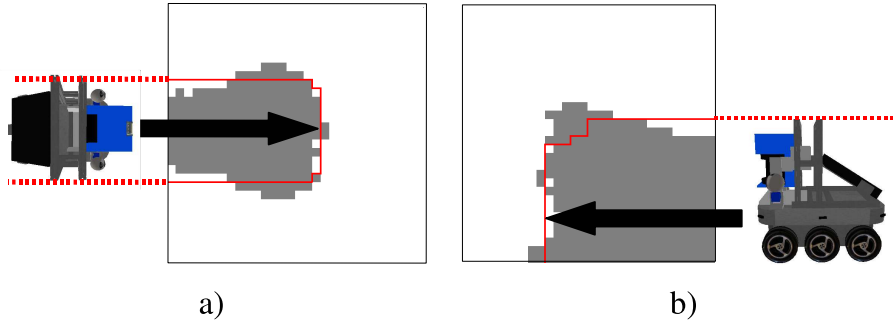


Figure 6: Prediction of collision boundaries and body dimensions in affordance perception. In (a) a box (illustrated by small gray/white squares) is shifted along longitudinal and lateral axes, and in (b) the box is shifted along longitudinal and vertical axes. Gray squares mark the positions where the robot predicted non-traversability. The lines that extend from the robot’s body correspond to actual collision boundaries for the go-forward behavior. The height, width and depth of these lines correspond to critical points for drive-under-ability, pass-through-ability, and go-forward-ability, respectively.

Specifically, 100 – 400 of the features were chosen to be relevant and the tolerance parameter was chosen as 250 – 500. This setting provides a prediction accuracy of approximately 87% in environments randomly generated.

The prediction accuracy is not higher since the traversability of an environment is a complex function of the individual objects as well as their layout in the environment. For instance, even small differences in the point of contact between the robot’s body and a lying cylinder can affect the outcome of the interaction. Moreover, due to line-of-sight some objects may be invisible. For instance, a lying cylinder can become non-traversable due to a box behind it. Finally, the grid square representation may lump patches from different objects, such as a patch from sphere and a patch from a box, producing confusing instances for the classifier.

We conducted experiments using both the simulated and the physical robot in settings inspired by the experimental settings used in Ecological Psychology to evaluate:

- whether the learned *affordances are relative* to the robot,
- whether the learning of *affordances provided perceptual economy* to the robot, and
- whether the learned *affordances generalized well in novel environments*, that were not interacted during training.

## 5.2 Are learned affordances relative?

The first set of experiments aimed to analyze whether the learned traversability affordances were related to the physical characteristics of the robot, such as its body dimensions and the capabilities of the robot.

### 5.2.1 Body size

The robot interacted with a random number of boxes that hang (fixed) in space within the cubic  $1\text{m}^3$  volume in front of the robot. The dimension [5cm – 20cm], position and orientation of the boxes as well as their number [0 – 12] are kept random. For each behavior, 160 relevant features were used to yield approximately 90% accuracy in predicting traversability.

In another set of experiments, a box is shifted along different axes with 3cm gaps and for each position the robot predicted the existence of traversability for the go-forward behavior. The results (Figure 6) show that the robot had acquired a “sense of its body” and was taking its body size into account in predicting traversability.

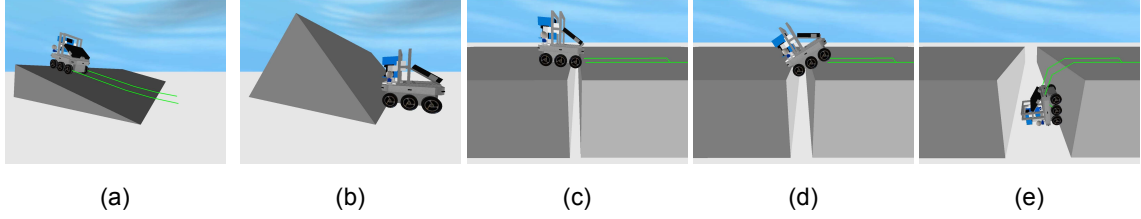


Figure 7: (a) *Climb-able* and (b) *not-climb-able* ramps. (c) *Cross-able* and (d,e) *not-cross-able* gaps.

Table 1: The predicted and actual *critical angles* of the ramps for climb-ability.

Ramp Orient.	Predicted/Actual <i>Critical Angles</i> for Behaviors		
	↑	↗	↘
-45°	12° / 11.5°	33.5° / 28°	- / -
-30°	10° / 8.5°	15.5° / 13.5°	30.5° / 26.5°
-15°	8° / 7.5°	9.5° / 9.5°	32° / 33.5°
0°	9° / 8°	7.5° / 8°	14° / 11°
15°	8° / 7.5°	8.5° / 8°	8.5° / 8.5°
30°	9° / 8.5°	8.5° / 8°	8° / 7.5°
45°	12° / 11.5°	9° / 8.5°	9° / 8.5°

### 5.2.2 Ramps

The robot interacted with ramps placed in its vicinity at random position, orientation, width, height and slopes (Figure 7). The most relevant 160 features were selected to yield approximately 95% prediction success. We compared the *critical slope* values of the ramp beyond which it becomes non-climb-able (corresponding to actual) or is perceived as non-climb-able (corresponding to predicted). Table 1 shows that the predicted and actual critical angles depend on the orientation of the ramp and behavior of the robot, and are very close.

### 5.2.3 Gaps

The robot interacted with gaps on the ground that are randomly placed within a distance of [10 cm - 100 cm] in front of the robot (Figure 7). The most relevant 160 features were used to yield approximately 95% accuracy. We analyzed the change in *critical width* for cross-ability affordance, and compared the actual and predicted values in Table 2. The results show that both the actual and predicted *critical widths* change with the relative orientation of the gap as compatible with the dynamics of the robot.

We argue that gap and ramp experiments are similar to Warren and Whang’s study (Warren &

Table 2: The predicted and actual *critical widths* of gaps for cross-ability.

Gap Orient.	Predicted/Actual <i>Critical Widths</i> for Behaviors		
	↑	↗	↘
-45°	30cm / 30cm	33cm / 33cm	36cm / 36cm
-30°	27cm / 30cm	27cm / 27cm	24cm / 24cm
-15°	21cm / 21cm	15cm / 18cm	15cm / 15cm
0°	12cm / 12cm	9cm / 18cm	15cm / 21cm
15°	24cm / 24cm	24cm / 24cm	21cm / 24cm
30°	30cm / 30cm	24cm / 27cm	21cm / 24cm
45°	30cm / 30cm	21cm / 27cm	27cm / 33cm

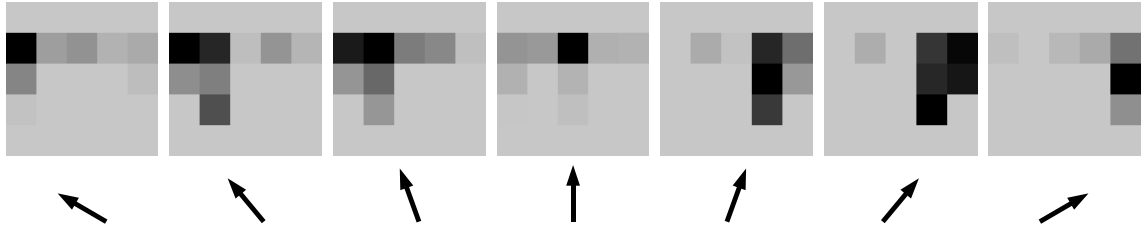


Figure 8: Grids relevant for different move behaviors.

Whang, 1987) on go-through-ability of apertures, Marcilly and Luyat’s study (Marcilly & Luyat, 2008) on pass-under-ability of barriers, Kinsella-Shaw et al.’s study (Kinsella-Shaw et al., 1992) on the walkability of slanted surfaces, and Jiang and Mark’s study (Jiang & Mark, 1994) on the cross-ability of gaps.






### 5.3 Do learned affordances provide perceptual economy?

We analyzed the number of relevant features chosen during the experiment described in Section 5.1 to show that at most 10% of the features was sufficient to predict traversability.

#### 5.3.1 Spatial distribution of relevant features

Figure 8 shows the grid squares that include relevant features for different behaviors where the darkness of a grid square is proportional to its relevancy. Couple of observations can be made. First, the bottom of the image corresponding to the robot’s body, and the top of the image which lies above the robot’s height were discovered to be irrelevant for all the behaviors. Second, the relevant grid squares tend to be aligned with the direction of the movement. Only the grid squares at the center of the range image are discovered to be relevant for the go-forward behavior, whereas the relevant grid squares for behaviors that turn left are grouped on the left part of the range image.

#### 5.3.2 Distribution of feature categories

The features were grouped into *i*) distance related ones, *ii*) shape related ones in lateral axis, and *iii*) shape related ones in longitudinal axis in order to analyze their relevancy. When the most relevant 20 features are considered, 65%, 30%, and 5% of them correspond to distance, lateral shape, and longitudinal shape related groups, respectively. Hence, the vertical shape of the objects is more important than their horizontal shape for perceiving traversability. Although  and  have horizontally different shapes, they have the same traversability affordance. On the other hand, the vertical shape distinguishes the traversabilities of , , and . Note however that if one considers all the most relevant 320 features, the number of features of shape relate groups becomes much larger than the distance related group since the shape related groups include more features compared to distance related group.

### 5.4 Do learned affordances generalize?

#### 5.4.1 Novel simple objects

We restricted the object types interacted during the exploration phase of the go-forward behavior and evaluated the traversability prediction accuracy for all object types (Table 3). In both phases, a single object with various positions, sizes and orientations were presented. The following observations can be made:

- When the training set includes only traversable objects (case 4), the classifier predicts traversability in all cases. When only non-traversable objects are included (cases 2 and

Table 3: Generalization of learned affordances

Case	Training obj. types	Prediction Accuracy (%)			
1		96	95	86	100
2		66	97	94	31
3		75	98	99	30
4		55	30	34	100
5		96	98	91	88
6		97	98	96	80
7		95	83	78	100
8		70	97	97	30
9		93	93	86	100
10		95	93	91	100
11		95	98	93	81
12		97	92	87	100
13		95	98	94	100
14		95	96	90	100
15		95	97	92	100

3), the traversability of the environment is mainly determined by the relative position of the object.

- In case 1, the robot is trained with only , yet it is able to predict the affordances of all other object types that it did not interact with before. This is due to the fact that a can be traversable or non-traversable, depending on its relative orientation with respect to the robot. The classifier correctly predicts that and are non-traversable, and that were traversable.
- We believe that the main reason behind the successful prediction of traversability for novel objects in cases 5, 6, 7, 11, 12, 13, and 15 is due to the inclusion of in training. The accuracy for is increased in some cases because other objects included into the training have similarities with .
- In case 9, since the training set contains samples for both success and fail, the affordances of novel objects ( and ) are also correctly predicted.

#### 5.4.2 Complex real-world objects

After training with only simple objects, as reported in Section 5.1, we evaluated the affordance prediction on three real-world objects shown in Figure 9. In first setup, the height and orientation of the wooden table was varied so that in some situations the robot can drive below the table and in some situations it cannot. The position and orientation of the iron table, and the height of the connecting bars were varied in the second setup. The structure of the chair was kept fixed in the third setup, but its position and orientation was varied. The prediction success rates were evaluated to be 85.5%, 81.7%, and 94.7%, where 64.9%, 74.2%, and 68.4% of the objects were traversable, respectively. The prediction accuracy of the iron table was lower compared to wooden one since the legs and bars were smaller in size.

### 5.5 Full demonstration

We proposed two execution architectures, namely *aggressive navigation* and *cautious navigation* (see Figure 10) in order to test the navigation in unstructured environments.

The aggressive navigation architecture minimizes the turns the robot has to make during navigation by prioritizing the move behaviors, as shown in Figure 10(a). Specifically, the architecture

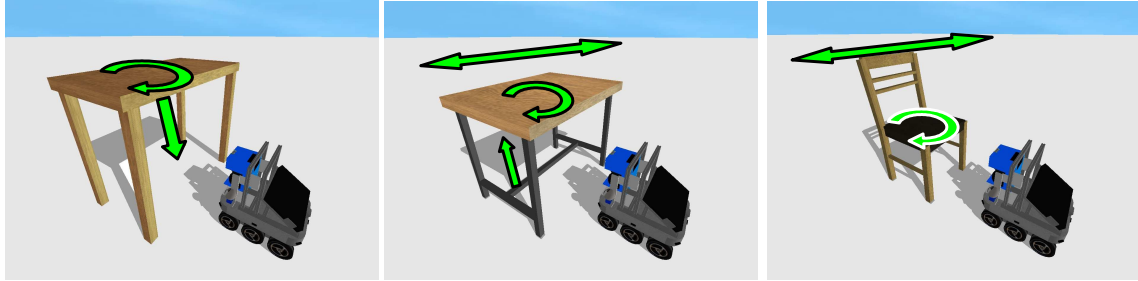


Figure 9: Left: The wooden table model. Middle: The iron table model. Right: The chair model.

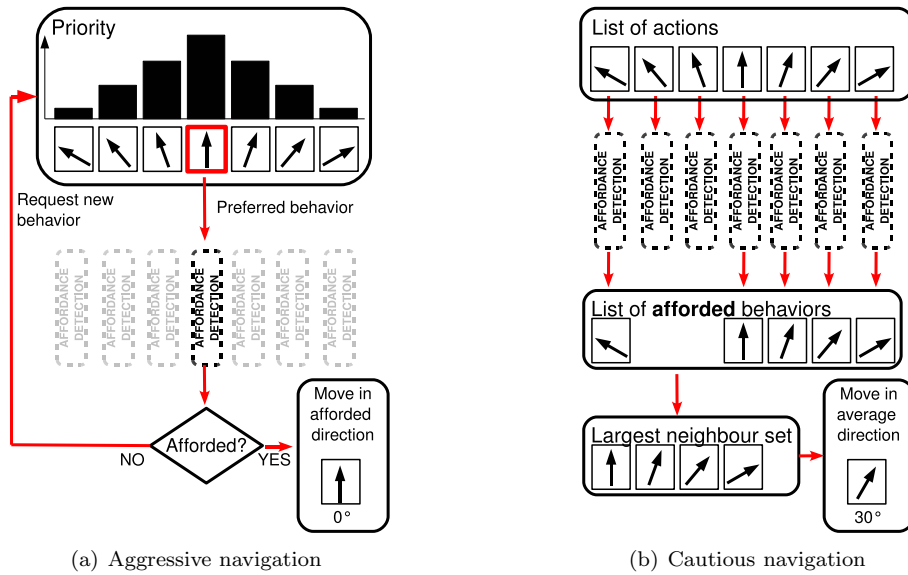


Figure 10: In aggressive navigation, the high-priority behavior is immediately executed if it is afforded. In cautious navigation, all afforded behaviors are considered and the average direction of neighbour afforded behaviors is used.



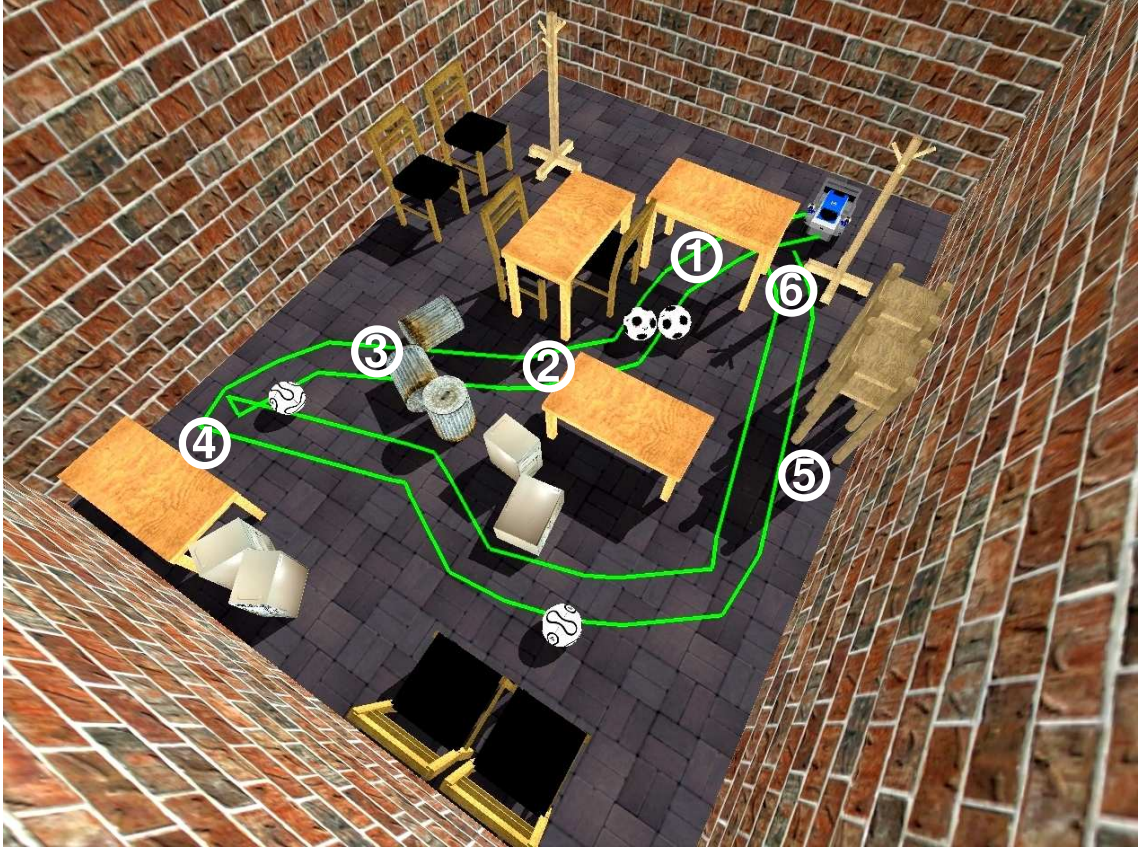


Figure 11: The path of the robot in a virtual room as driven by the aggressive navigation architecture.

queries the existence of affordances for each behavior in the order of priority and executes the first one afforded.

We used the classifiers trained through interactions with simple objects, as reported in Section 5.1, in a virtual room cluttered with office objects of different sizes and types (Figure 11). Situation (1) is predicted to be traversable for the go-forward behavior since the table is high enough to drive-under, and the width between the legs is wide enough to pass through. Situations (4) and (5) are not traversable for the go-forward since the coffee table is not sufficiently high and the aperture between the legs of the shelf is narrow. The robot is able to pass-through the legs of different tables in situation (2) and correctly predicts the traversability of the garbage bins in situation (3). The robot makes an incorrect prediction in the last step (6), i.e. it predicts that the aperture width between the leg of the table and extension of the hanger affords traversability.

The cautious navigation architecture, sketched in Figure 10(b), takes a more conservative approach in order to minimize the risk of collisions. Different from aggressive navigation, the robot moves only if more than one neighbor behavior is afforded, and heads in the average direction of the largest neighbor set.

The cautious navigation architecture is used to drive the robot in Figure 12. The range images of a number of situations marked in this figure are shown in Figure 13 together with the afforded behaviors. Figure 13(a) shows that the robot correctly learned its own body dimensions, i.e. the *go-under-ability* of the table and *pass-through-ability* through its legs are correctly perceived. The largest set of afforded neighbour behaviors in this case is  $\{\uparrow, \nearrow\}$ , so the robot first rotates  $10^\circ$  around itself and then goes forward.

In Figures 13(b), the robot goes over the cylindrical object whose orientation makes it convenient to push, and rolls it aside. Moreover, the robot deals with the confusing situation where the



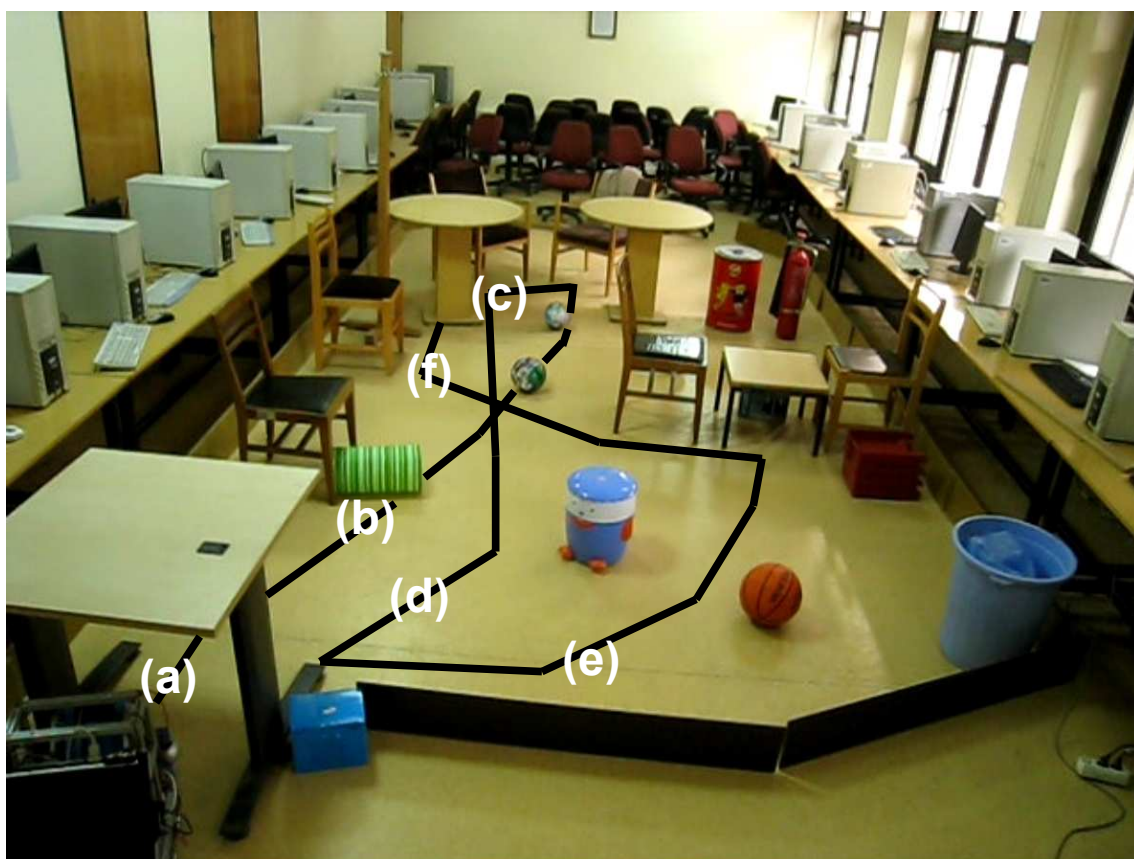


Figure 12: The approximate course of the robot resulting from the execution of the cautious controller described in Figure 10(b) in a real room. Movie available at <http://kovan.ceng.metu.edu.tr/traversability>.

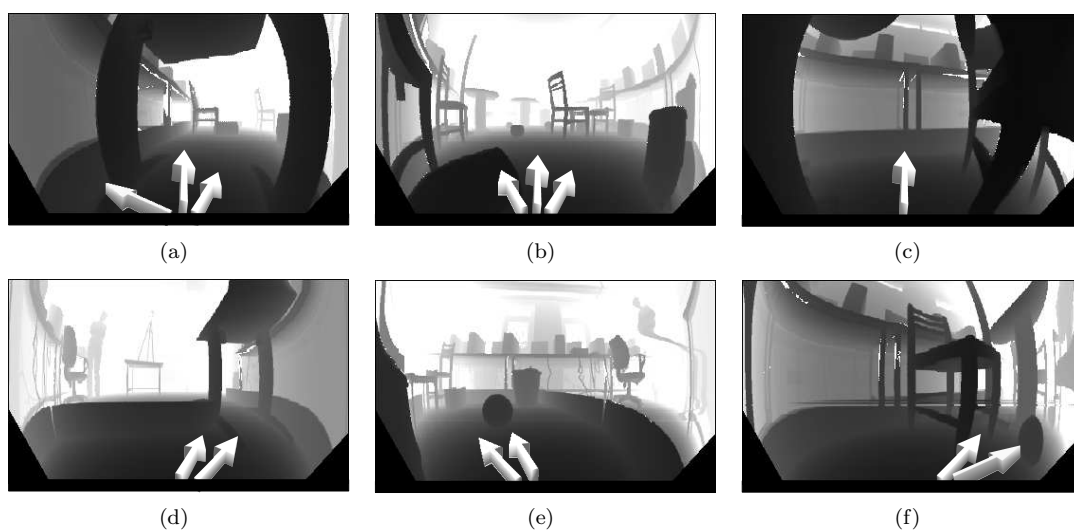


Figure 13: Range images that are used to decide the traversability of the environment for some situations identified in Figure 12.

traversable cylinder and the non-traversable chair locate in the same direction; it decides that the behaviors  $\nearrow$  and  $\nwarrow$  are not afforded. In (e), although the ball is perceived to be traversable, the robot does not go towards it since the other afforded behavior moves the robot further to the left than expected.

In Figure 13(c) only go-forward ( $\uparrow$ ) behavior is afforded, however the robot does not drive forward since at least two neighbouring behaviors should be afforded. Indeed the aperture width between the table base on the left and the chair on the right is not large enough for the robot to pass-through as predicted. This example situation shows how the robot is protected from a collision by being *cautious*. However in Figures 13(d) and 13(f), the robot cannot avoid collision even in this mode because the parts of the table and chair are very small compared to training objects.

## 6 Related work

### 6.1 Affordances

The concept of affordances has received interest from the robotics community and there has been a rise in the number of studies that aim to use this elusive concept at different levels of robot control, as reviewed in detail in our earlier work (Şahin et al., 2007). Most of these studies however are concerned with the manipulation of objects with robotic arms.

In the context of manipulation based affordance learning, (Fitzpatrick, Metta, Natale, Rao, & Sandini, 2003) studied the rollability affordances of the objects using vision, and claimed that manipulation can be used to ease and ground visual perception (Fitzpatrick & Metta, 2003). Although the rolling direction of objects is learned in these studies, the objects in the environment were differentiated using their colors only, and the discovery of the distinctive features relevant for rollability was not addressed. (Montesano, Lopes, Bernardino, & Santos-Victor, 2008) proposed a general probabilistic model based on Bayesian networks to learn the relationship between actions, objects, and effects through interaction with environment, however the object classes formed are not based on the generated effects. Tool affordances for a robot are learned in (Sinapov & Stoytchev, 2008) but the the object dealt with is kept fixed, so affordances of the objects are not learned. In (S. Griffith, Miller, & Stoytchev, 2009) the object affordances are learned through interaction for a task that requires categorization of container and non-container objects. In (Montesano et al., 2008), a Bayesian network is used to learn object affordances, i.e. the structural relations between object properties, actions and generated effects. The object properties that have no influence on other components of the system can be discovered by the network and filtered out during task execution.

The concept of ‘object-action complexes’ (OACs), which argues that objects and actions are tightly linked, is also relevant to affordances. Along the lines of the concept of OACs, the work by (Kraft et al., 2008) uses the assumption that combinations of certain visual features suggest certain grasps of ‘things’ in a scene, and names an object the set of visual features that move in the scene in accordance with the executing grasping action. This work is extended in (Petrick et al., 2008) by learning effects of actions (such as filling, moving) from its preconditions and its effects (using kernel perceptron learning). In (Wörgötter, Agostini, Krüger, Shylo, & Porr, 2009), the concept of OACs is linked to the predictability of the environment and the body of the robot and how these can be used to improve the robots model of the world and itself.

### 6.2 Traversability

The interest in the learning of traversability has recently been fueled by the LAGR (Learning Applied to Ground Robots) program (Jackel, Krotkov, Perschbacher, Pippine, & Sullivan, 2007). In this program, the robots are required to learn the traversability characteristics of the environment and plan paths based on short-range and long-range traversability predictions. The traversability is not defined simply as obstacle avoidance, and the robots are expected to avoid from bushes and shrubs while driving over soft grasses of similar height.

Most teams that compete in this program (Bajracharya, Howard, Matthies, Tang, & Turmon, 2008; Shneier et al., 2008) carry out training under the self-supervision of the robot’s own signals, such as bumpers, inertial navigational system, wheel encoders and camera images. The robot learns and predicts short-range traversability of the environment mainly through range images obtained from laser rangefinder or stereo vision; and long-range traversability through color camera features. First, the pixels of range and color images are projected onto local grid squares and learning is performed in the constructed map environment. In this approach, regions in the local map are marked as traversable or not.

In (Ollis, Huang, Happold, & Stancil, 2008) the robot plans paths directly over images obtained from color cameras, however their intermediate steps include high-level 3D processing such as ground plane detection, identification of points above the ground plane, and terrain slope computation. In some studies, perception of obstacles and free-spaces are hand-coded and learning is done based on the obstacle-ground distribution of these objects (Shneier et al., 2008). In others, such as (Bajracharya et al., 2008), features related to the physical affordances are carefully identified by hand and used directly. Although in (Kim, Sun, Oh, Rehg, & Bobick, 2006), the authors relate traversability to Gibsonian affordances, and claim to “learn a direct mapping from observations to affordances”, we believe that the use of global maps is orthogonal to the Gibsonian view of perception.

### 6.3 Robot learning

Relevant regions/cues/features in the environment are automatically discovered in many robotic tasks such as robot localization (Zhang, Lihua, & Adams, 2005; Li & Kosecka, 2006), object tracking (Jung & Sukhatme, 2004), and robot navigation (Peng & Peters, 2005; Bur, Tapus, Ouerhani, Siegwar, & Hiigli, 2006). In (Zhang et al., 2005) the features used for simultaneous localization and map building are filtered out to increase the speed of the process. The features that increase uncertainty in robot localization are filtered out using an entropy-based method. Similarly (Li & Kosecka, 2006) selected the most discriminative features to recognize the location of the robot by measuring the information entropy that is calculated from posterior probabilities of location classes given the feature values. (Jung & Sukhatme, 2004) selected a number of image points among many of the detected ones to track the moving objects from a mobile platform. In (Peng & Peters, 2005), the robot is tele-operated first, discrete motor states are differentiated and the salient features that consistently co-occur in same motor states are discovered and later used autonomous navigation phase. In navigation of the robot, (Bur et al., 2006) selected and used the features that are persistent over the course of previous runs. These studies used feature selection as a means to discover features that will allow the recognition of ‘visual landmarks’ and did not use them to learn general relations about the environment. In (Floreano, Kato, Marocco, & Sauser, 2004) the action-relevant features are discovered through evolutionary algorithms and later used in robot navigation, in a similar vein to our work.





















## 7 Discussion

We used two off-the-shelf methods from Machine Learning research; namely the ReliefF method to extract relevant features for traversability, and the SVM’s as classifiers for predicting traversability. These specific methods were chosen over other alternatives since they scale well with the dimension of the inputs, with the size of the training dataset, and perform robustly when faced with noisy data. Moreover, after training, the SVMs store only a small number of parameters and have low computational complexity during execution.

### 7.1 Feature Selection

The methods that select relevant features can be roughly categorized in two groups; namely *wrappers* and *filters* (Blum, 1997). The filter methods select features based on metrics such as distance and correlation without considering how the selection would affect the performance in classifier phase. The wrapper methods on the other hand measure the relevance of features based on the

Table 4: The 20 most relevant features discovered by the ReliefF method for the go-forward behavior.

Rank	Type	Feature	Grid position
1	distance	min	
2	latitude( $\theta$ )	$[0^\circ, 20^\circ]$	
3	distance	min	
4	latitude( $\theta$ )	$[0^\circ, 20^\circ]$	
5	latitude( $\theta$ )	$[60^\circ, 80^\circ]$	
6	distance	min	
7	distance	mean	
8	distance	min	
9	latitude( $\varphi$ )	$[0^\circ, 20^\circ]$	
10	distance	mean	
11	distance	min	
12	distance	min	
13	distance	min	
14	latitude( $\varphi$ )	$[60^\circ, 80^\circ]$	
15	distance	mean	
16	distance	mean	
17	longitude( $\varphi$ )	$[0^\circ, 20^\circ]$	
18	distance	min	
19	distance	min	
20	latitude( $\theta$ )	$[60^\circ, 80^\circ]$	





















performance of classifier used in the second phase and can produce near-optimal results (Dash & Liu, 1997). Although wrapper methods give better results compared to filter methods in general, the classification phase tends to have high computational complexity in most applications and does not scale well with the dimension of the inputs and the size of the training dataset. Filter methods have lower computational complexity since they do not utilize the classification phase in computing the relevancy of the features. However, they tend to produce less than optimum sets.

Table 4 lists the 20 most relevant features discovered by the ReliefF method for predicting the traversability of the go-forward behavior. The feature list can roughly be categorized into three sets: (1) features measure the minimum or the mean of the distance values coming from the central grid squares, and (2) the latitude and (3) longitude features coming from almost similar grid squares measuring the normal vector histograms between certain degrees. In order to analyze the redundancy of these features, we used the *sequentialfs* (sequential feature selection) method provided in the MATLAB package<sup>4</sup>, within the wrapper category. The *sequentialfs* method generates near-optimal relevant feature sets in a way similar to the one used in Schemata Search (Moore & Lee, 1994). Starting from an empty relevance feature set, it selects one feature and adds it to the feature set of previous iteration. At each iteration, a candidate feature set for each not-yet-selected feature is formed by adding the corresponding feature to the previous feature set. Then, the candidate feature sets are evaluated through 10-fold cross-validations on SVM classifiers that use these candidate feature sets. The best performing candidate set is then transferred to the next iteration.

Table 5 ranks the most important 20 features found after the 20 iterations of *sequentialfs* method. The table also includes the rank of each feature as evaluated by the ReliefF method for comparison purposes. As expected, the most relevant feature discovered by the *sequentialfs* method ranks also high on the ReliefF ranking. It can be seen that due to the incremental nature of the *sequentialfs* method, there is little correspondence between the rankings of the two methods. However, given that *sequentialfs* method produces a more optimal set of features, we can now go back to the set of relevant features listed in Table 4 to support our claim that there exist a lot of redundancy among the information carried out by the features that are ranked high by the ReliefF

<sup>4</sup><http://www.mathworks.com/access/helpdesk/help/toolbox/stats/sequentialfs.html>

Table 5: The 20 most relevant features discovered by the *sequentialfs* method for the go-forward behavior.

<i>sequentialfs</i> rank	Type	Feature	Grid position	RelieFF rank
1	distance	min		3
2	latitude( $\theta$ )	$[40^\circ, 60^\circ]$		113
3	latitude( $\theta$ )	$[20^\circ, 40^\circ]$		70
4	latitude( $\theta$ )	$[-40^\circ, -20^\circ]$		428
5	latitude( $\theta$ )	$[20^\circ, 40^\circ]$		243
6	longitude( $\varphi$ )	$[40^\circ, 60^\circ]$		423
7	longitude( $\varphi$ )	$[40^\circ, 60^\circ]$		290
8	distance	min		438
9	longitude( $\varphi$ )	$[0^\circ, 20^\circ]$		302
10	latitude( $\theta$ )	$[60^\circ, 80^\circ]$		14
11	longitude( $\varphi$ )	$[-80^\circ, -60^\circ]$		256
12	latitude( $\theta$ )	$[60^\circ, 80^\circ]$		291
13	latitude( $\theta$ )	$[-20^\circ, 0^\circ]$		280
14	longitude( $\varphi$ )	$[40^\circ, 60^\circ]$		585
15	distance	min		167
16	longitude( $\varphi$ )	$[0^\circ, 20^\circ]$		293
17	distance	mean		232
18	latitude( $\theta$ )	$[20^\circ, 40^\circ]$		196
19	longitude( $\varphi$ )	$[0^\circ, 20^\circ]$		131
20	distance	min		46

method. For instance, only 5 distance related features appear in Table 5, as opposed to 13 in the Table 4 indicating the redundancy in the relevant feature set.

Figure 14 plots the performance of the classifiers that are trained with the most important  $n$  features obtained by the *sequentialfs* ( $n \leq 20$ ) and the RelieFF method ( $n \leq 320$ ) for predicting the traversability of the go-forward behavior. The evaluation was made in environments with the same characteristics of the training environment where varying number of different types of objects are randomly placed in the frontal area of the robot. As shown, although RelieFF method has better performance at the end; for a given number of relevant features, *sequentialfs* method performs better than RelieFF.

## 7.2 Traversability problem

The difficulty inherent in learning traversability of the robot, as studied in this paper, begs further analysis in order to ensure that the problem is not reduced to a trivial one through the choice of the particular feature representation. Figure 14 shows that classifiers can achieve prediction performance of approximately 64% (72%) using the most relevant feature (minimum distance from one of the central grid squares) discovered by the RelieFF (*sequentialfs*) method. The inclusion of the next three features raises the performance to 78% (80% for *sequentialfs*), and the performance gradually reaches 87% with the use of 320 features.

In order to analyze in detail how the inclusion of additional features contributed to the prediction performance, we used seven exemplary setups as shown in Table 6. It can be seen that through the use of the most relevant feature only, classifiers merely link the existence/non-existence of an object in the frontal area to traversability and do not take into account their rollability. The inclusion of the second most relevant feature detected by the *sequentialfs* method allows the detection of traversability in lying cylinders that are properly aligned but fails the detection of traversability in spheres. In a similar manner, the inclusion of the second most relevant feature detected by RelieFF method allows the classifier to detect the traversability of spheres but fails on lying cylinders. As can be seen, the classifiers trained by the most three (four) relevant features detected by the *sequentialfs* (RelieFF) method are able to detect the traversability affordances in the first six setup that included single objects. However, when the scene is cluttered with multiple objects,

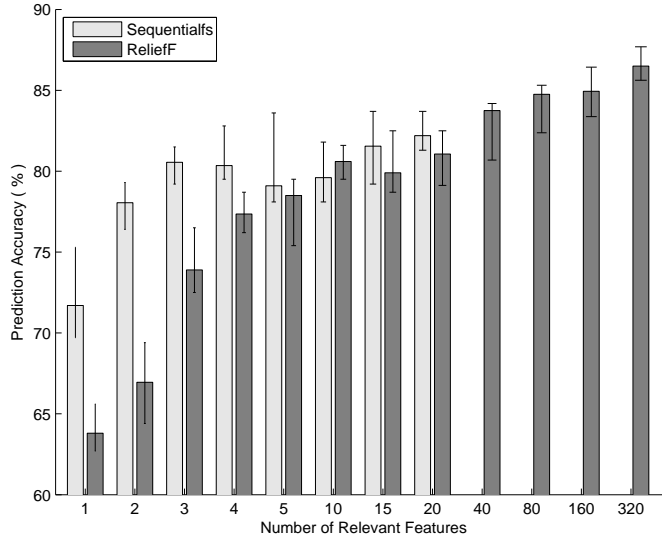


Figure 14: The prediction accuracies obtained with the most relevant  $n$  features in training-like test environments.

such as the seventh setup, where the robot faces both a close-by sphere and a further-away box in its frontal view, then the prediction becomes difficult, and requires the use of 320 features.

Table 6: The traversability prediction results of classifiers in eight exemplary setups.  $Seqfs(n)$  and  $ReliefF(n)$  denote classifiers trained with the first  $n$  relevant features discovered by the *sequentialfs* and *ReliefF* methods.

	1	2	3	4	5	6	7	8
Seqfs (1)	✓	✓	✓	X	✓	X	✓	X
Seqfs (2)	✓	✓	✓	X	✓	✓	✓	X
Seqfs (3)	✓	✓	✓	✓	✓	✓	✓	X
Seqfs (4)	✓	✓	✓	✓	✓	✓	✓	X
Seqfs (20)	✓	✓	✓	✓	✓	✓	✓	X
ReliefF (1)	✓	✓	✓	X	✓	X	✓	X
ReliefF (2)	✓	✓	✓	✓	✓	X	✓	X
ReliefF (3)	✓	✓	✓	X	✓	X	✓	X
ReliefF (4)	✓	✓	✓	✓	✓	✓	✓	X
ReliefF (320)	✓	✓	✓	✓	✓	✓	✓	✓

A closer inspection of Figure 14 shows that as the number of features used by the classifier increases from 20 to 320, the performance merely increases from 82% to 87%. Hence, one may question whether the amount of performance gain is sufficient to justify the extra cost. At this point, we would like to point out that the prediction performances reported in the above experiments are determined by the distribution of the test environments, and may not be very representative of the real-world performance to be expected from the robot. Towards this end, we conducted an experiment to evaluate the performance of the classifiers by hanging boxes (which are non-traversable) around the *critical points* for traversability. Specifically, in a similar setup shown in Figure 6, we systematically placed boxes (non-traversable) inside or outside of the collision boundaries within a 10cm band as shown in Figure 15(a) and plotted the prediction performance in Figure 15(b). The results show that as the number of relevant features increase from 20 to 320, the performance

increases from 65% to 85%, a significant gain in borderline situations.

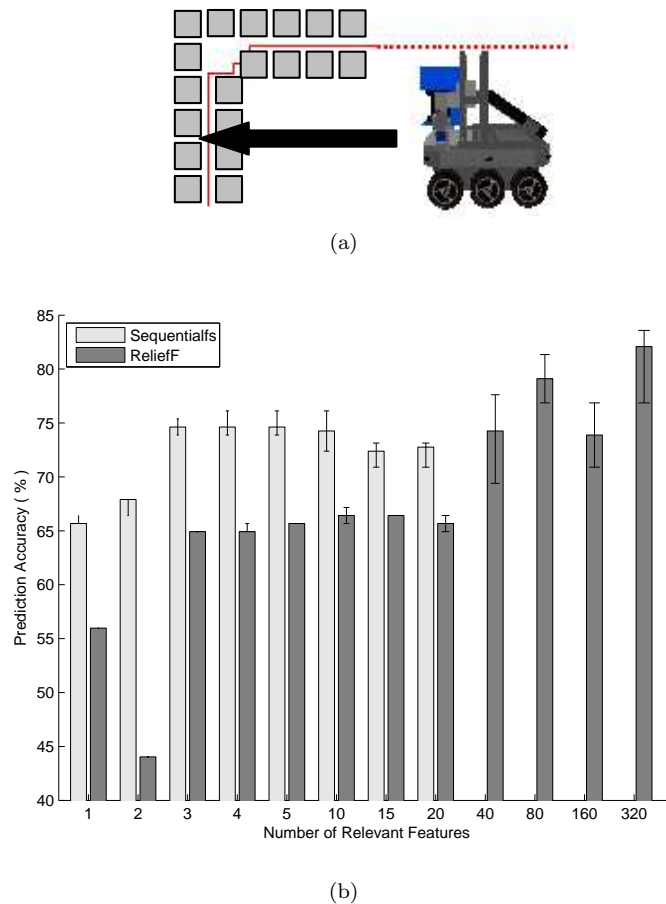


Figure 15: (a) Boundary experiment setup. The robot is asked to predict the traversability of environments which include a hanging box that is placed either on one of the given gray squares or a position in between these squares. (b) The prediction accuracies obtained with the most relevant  $n$  features in the boundary experiment.

Finally, we would like to point out that the sufficiency of using a linear kernel in SVM classifiers does not necessarily imply the simplicity of the problem, since many learning problems that require complex high-dimensional kernels at low-dimensional feature spaces are transformed into simpler problems that can be linearly separable through the use of high-dimensional features.

## 8 Conclusion

In this paper, we studied the learning and perception of traversability affordance in organisms and robots with the hope of appealing to readers from both Ecological Psychology and Autonomous Robotics. Hence the contributions of this paper are two-fold: first, from a robotics point of view, it presents a method for the learning and perception of traversability on mobile robots using range images. Specifically, we proposed a set of features for representing the shape and distance information on range images that are shown to provide a good degree of generalization, and a scalable method towards learning affordance relations. The learning method uses off-the-shelf machine learning methods that are highly scalable with the input dimension. The proposed method shows that one can start with a large feature vector that contains all types of feature

detectors that one can propose, and have it reduced down to only a fraction after training. In this sense, the robot can minimize the load on its perceptual processing after learning to achieve perceptual economy. A systematic analysis of the method and its performance under different parameter settings, and in both simulated and physical environments, showed that despite the simplicity of perceptual representation, the method can provide a good degree of generalization, as demonstrated in Section 5.5 where upon training with only simple object types in a simulated environment, the robot can navigate successfully among complex objects in the real-world.

Second, from an Ecological Psychology point of view, the paper shows that the formalization proposed in our earlier work (Şahin et al., 2007), can indeed be used to make the robots learn the affordances in its environment. Through experiments that are inspired by the ones used in Ecological Psychology, we show that three of the main attributes that are commonly associated with affordances (namely, affordances being relative, providing perceptual economy and providing general information) are simply a consequence of learning from the interactions of the robot with the environment.

The study presented in this paper has a number of limitations that can mainly be attributed to the use of our 3D range sensing equipment, which takes almost 40 seconds to produce a range image. First, the speed of sensing limits the reactivity of the robot, and doesn't leave much room for the robot to immediately perceive and react to changes in the environment. Second, the slowness also makes it prohibitive to obtain large quantities of data to be used for learning, which was tackled through the use of a physical simulator. However, the physical simulators bring in their own constraints, such as the difficulty of access to 3D models of real-world objects which then limits the type of interactions that can be explored in simulation. These limitations can be addressed through using stereo vision systems that operate using standard cameras or through the use of 3D cameras that can provide range images at large frame rates.

Finally, we would like to point out that although the use of range images makes it easier to link and generalize the perceptual features with the physical affordances of the environment, the proposed methodology does not pose any limitation on the type sensing device. As a matter of fact, the use of regular camera images may indeed be used to discover/develop image features that are relevant to affordances may produce interesting insights to computer vision similar to the ones shown in (Fitzpatrick & Metta, 2003).

## Acknowledgements

This work was partially funded by the ROSSI project (FP7-ICT-21625). Emre Ugur acknowledges the financial support of TÜBİTAK through National Scholarship Programme for Ph.D. Students.



## References

- Adolph, K. E., Bertenthal, B. I., Boker, S. M., Goldfield, E. C., & Gibson, E. J. (1997). Learning in the development of infant locomotion. *Monographs of the Society for Research in Child Development*, 62(3).
- Bajracharya, M., Howard, A., Matthies, L. H., Tang, B., & Turmon, M. (2008). Autonomous off-road navigation with end-to-end learning for the lagr program. *Journal of Field Robotics*, 26(1), 3–25.
- Blum, A. (1997, December). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97, 245–271.
- Bur, A., Tapus, A., Ouerhani, N., Siegwar, R., & Hiigli, H. (2006). Robot navigation by panoramic vision and attention guided features. In Y.Y.Tang, S.P.Wang, G. Lorette, D. Yeung, & H.Yan (Eds.), *Proceedings of the 18th international conference on pattern recognition* (pp. 695–698). Hong Kong: IEEE.
- Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology*, 15(2), 181-195.
- Chemero, A., Klein, C., & Cordeiro, W. (2003). Events as changes in the layout of affordances. *Ecological Psychology*, 15(1), 19-28.
- Collett, T. (1977). Stereopsis in toads. *Nature*, 267, 349-351.
- Cornus, S., Montagne, G., & Laurent, M. (1999). Perception of a stepping-across affordance. *Ecological Psychology*, 11(4), 249-267.
- Şahin, E., Çakmak, M., Doğar, M. R., Uğur, E., & Üçoluk, G. (2007). To afford or not to afford: A new formalization of affordances toward affordance-based robot control. *Adaptive Behavior*, 15(4), 447–472.
- Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, 1, 131–156.
- Fitzpatrick, P., & Metta, G. (2003). Grounding vision through experimental manipulation. *Philosophical Transactions of the Royal Society: Mathematical, Physical, and Engineering Sciences*, 361(1811), 2165–2185.
- Fitzpatrick, P., Metta, G., Natale, L., Rao, A., & Sandini, G. (2003). Learning about objects through action -initial steps towards artificial cognition. In *Ieee international conference on robotics and automation* (p. 3140-3145). IEEE.
- Floreano, D., Kato, T., Marocco, D., & Sauser, E. (2004, March). Coevolution of active vision and feature selection. *Biological cybernetics*, 90, 218–28.
- Gibson, E., & R.D.Walk. (1960). The visual cliff. *Scientific American*, 202, 64–71.
- Gibson, E. J. (2000). Perceptual learning in development: Some basic concepts. *Ecological Psychology*, 12(4), 295-302.
- Gibson, E. J. (2003). The world is so full of a number of things: On specification and perceptual learning. *Ecological Psychology*, 15(4), 283-288.
- Gibson, E. J., Riccio, G., Schmuckler, M. A., Stoffregen, T. A., Rosenberg, D., & Taromina, J. (1987). Detection of the traversability of surfaces by crawling and walking infants. *Journal of Experimental Psychology*, 13(4), 533-544.
- Gibson, J. J. (1986). *The ecological approach to visual perception*. Lawrence Erlbaum Associates.
- Ingle, D., & Cook, J. (1977). The effects of viewing distance upon size preference of frogs for prey. *Vision Research*, 17, 1009-1019.
- Jackel, L. D., Krotkov, E., Perschbacher, M., Pippine, J., & Sullivan, C. (2007). The darpa lagr program: Goals, challenges, methodology, and phase i results. *Journal of Field Robotics*, 23(11–12), 945 – 973.
- Jiang, Y., & Mark, L. (1994). The effect of gap depth on the perception of whether a gap is crossable. *Perception and Psychophysics*, 56(6), 691–700.
- Jung, B., & Sukhatme, G. (2004). Detecting moving objects using a single camera on a mobile robot in an outdoor environment. In F. Groen, N. Amato, A. Bonarini, E. Yoshida, & B. Krose (Eds.), *Proceedings of the 8th international conference on intelligent autonomous systems* (pp. 980–987).
- Kim, D., Sun, J., Oh, S. M., Rehg, J. M., & Bobick, A. (2006, May). Traversability classification using unsupervised on-line visual learning for outdoor robot navigation. In *Ieee international conference on robotics and automation* (pp. 518–525). Orlando, FL.

- Kinsella-Shaw, J. M., Shaw, B., & Turvey, M. T. (1992). Perceiving walk-on-able slopes. *Ecological Psychology*, 4(4), 223-239.
- Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In D. Sleeman & P. Edwards (Eds.), *Proceedings of the 9th international workshop on machine learning* (pp. 249–256). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Kononenko, I., Simec, E., & Robnik-Sikonja, M. (1997). Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence*, 7(1), 39–55.
- Kraft, D., Pugeault, N., Başeski, E., Popovic, M., Kragic, D., Kalkan, S., et al. (2008). Birth of the object: Detection of objectness and extraction of object shape through object-action complexes. *International Journal of Humanoid Robotics*, 5(2), 247–265.
- Li, F., & Kosecka, J. (2006). Probabilistic location recognition using reduced feature set. In *Ieee international conference on robotics and automation* (pp. 3405–3410). IEEE.
- Lock, A., & Collett, T. (1979). A toads devious approach to its prey: A study of some complex uses of depth vision. *Journal of Comparative Physiology*, 131, 179-189.
- Marcilly, R., & Luyat, M. (2008). The role of eye height in judgment of an affordance of passage under a barrier. *Current Psychology Letters: Behavior, Brain and Cognition*, 24(1).
- Montesano, L., Lopes, M., Bernardino, A., & Santos-Victor, J. (2008). Learning object affordances: From sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, 24(1), 15–26.
- Moore, A., & Lee, M. (1994). Efficient algorithms for minimizing cross validation error. In R. Greiner & D. Schuurmans (Eds.), *Proceedings of the 11th international conference on machine learning* (p. 190198). Morgan Kaufmann.
- Ollis, M., Huang, W. H., Happold, M., & Stancil, B. A. (2008). Image-based path planning for outdoor mobile robots. In S. Hutchinson (Ed.), *Ieee international conference on robotics and automation* (pp. 2723–2728). California: IEEE.
- Peng, J., & Peters, A. (2005). Extraction of salient features for mobile robot navigation via teleoperation. In *Proceedings of the 24th american control conference* (Vol. 7, pp. 4903–4908). IEEE.
- Petrick, R., Kraft, D., Mourão, K., Pugeault, N., Krüger, N., & Steedman, M. (2008, July). Representation and integration: Combining robot control, high-level planning, and action learning. In P. Doherty, G. Lakemeyer, & A. Pobil (Eds.), *Proceedings of the 6th international cognitive robotics workshop* (pp. 32–41).
- Sedgwick, H. A. (2001). Visual space perception. In E. B. Goldstein (Ed.), *Blackwell handbook of perception* (chap. 5). Oxford: Wiley-Blackwell.
- S. Griffith, J. S., Miller, M., & Stoytchev, A. (2009). Toward interactive learning of object categories by a robot: A case study with container and non-container objects. In *Proceedings of the ieee 8th international conference on development and learning* (pp. 1–6). Shanghai, China: IEEE.
- Shneier, M., Chang, T., Hong, T., Shackelford, W., Bostelman, R., & Albus, J. S. (2008). Learning traversability models for autonomous mobile vehicles. *Autonomous Robots*, 24(1), 69–86.
- Sinapov, J., & Stoytchev, A. (2008). Detecting the functional similarities between tools using a hierarchical representation of outcomes. In *Proceedings of the ieee 7th international conference on development and learning* (pp. 91–96). IEEE.
- Szokolszky, A. (2003). An interview with Eleanor Gibson. *Ecological Psychology*, 15(4), 271-281.
- Uğur, E., Doğar, M. R., Soysal, O., Çakmak, M., & Şahin, E. (2006). *MACSim: Physics-based simulation of the KURT3D robot platform for studying affordances* (Tech. Rep.). (MACS Project Deliverable)
- Uğur, E., Oztop, E., & Şahin, E. (2009). Affordance learning from range data for multi-step planning. In *Proceedings of the 9th international conference on epigenetic robotics* (Vol. 146, p. 177-184). Lund University Cognitive Studies.
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley-Interscience.
- Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology*, 105(5), 683-703.
- Warren, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: body-scaled information for affordances. *Journal of Experimental Psychology*, 13(3), 371-383.

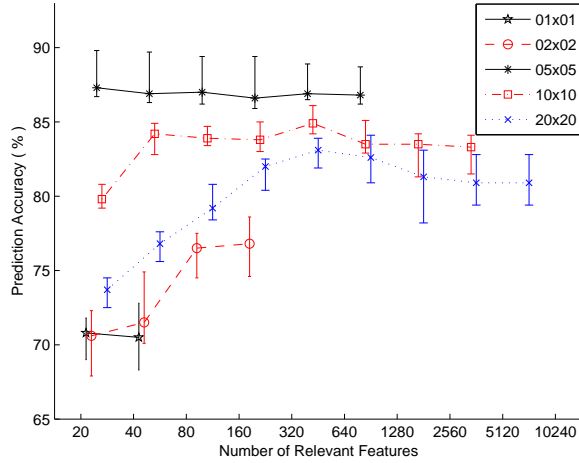


Figure 16: The prediction accuracies of the classifiers that are computed using different resolutions and number of relevant features for the go-forward behavior. The results of best, median, and worst classifiers are shown using bars.

Wörgötter, F., Agostini, A., Krüger, N., Shylo, N., & Porr, B. (2009, April). Cognitive agents : a procedural perspective relying on the predictability of object-action-complexes (oacs). *Robotics and Autonomous Systems*, 57(4), 420–432.

Zhang, S., Lihua, X., & Adams, M. D. (2005). Entropy based feature selection scheme for real time simultaneous localization and map building. In *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1175–1180). IEEE.

## A Parameter optimization for affordance prediction

We analyzed the parameters of the perception and learning system that affect affordance prediction performance. Specifically we optimized the values of the following parameters:

- The resolution of perceptual representation, denoted by  $r$ .
- The number of relevant features used in affordance perception, denoted by  $n$ .
- The cost parameter of SVM classifiers, denoted by  $c$ .

First we analyzed the effect of grid-resolution and number of relevant features on prediction performance. For this purpose, a separate classifier is trained for each grid-resolutions and each number of relevant features. Figure 16 shows the affordance prediction accuracies of different classifiers trained for body-size experiments using (Section 5.2.1), where medium resolutions (5x5 or 10x10) are found to give the best prediction accuracies.

However, the prediction accuracies of different move behaviors for different resolutions (see Figure 17(a)) indicate that there exists no single best resolution for all behaviors. A better and more invariant prediction performance across different behaviors can be achieved through using *overlapped* layers of grids (shown as the rightmost set of boxes in Figure 17(a)) for a 5x5 grid resolution. The effect of overlapped representation becomes even more evident when different types of objects are included in training as can be seen in Figure 17(b).

The relevant features that give best traversability accuracies differ for different move behaviors and for different environment types. If all objects are included into the environment, the shape of the objects play major role in traversability perception so more features should be used as shown in Figure 18.

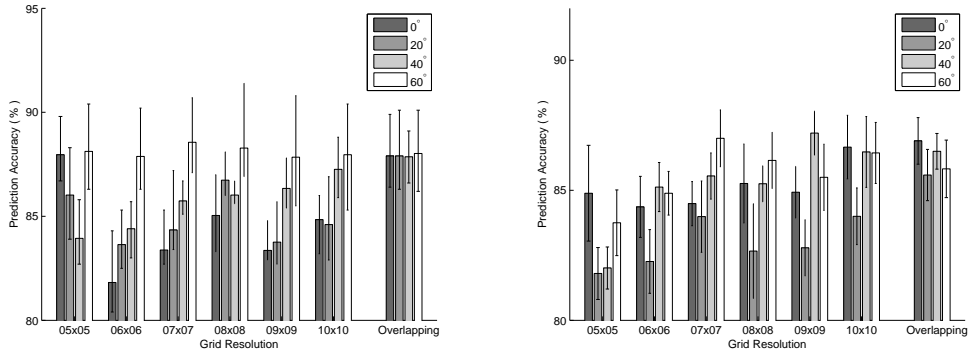


Figure 17: The prediction accuracies of different move behaviors for different resolutions. The cost parameter and number of relevant features are optimized to obtain best accuracy for each resolution and behavior. Only boxes are included in (a). All simple object are included in (b).

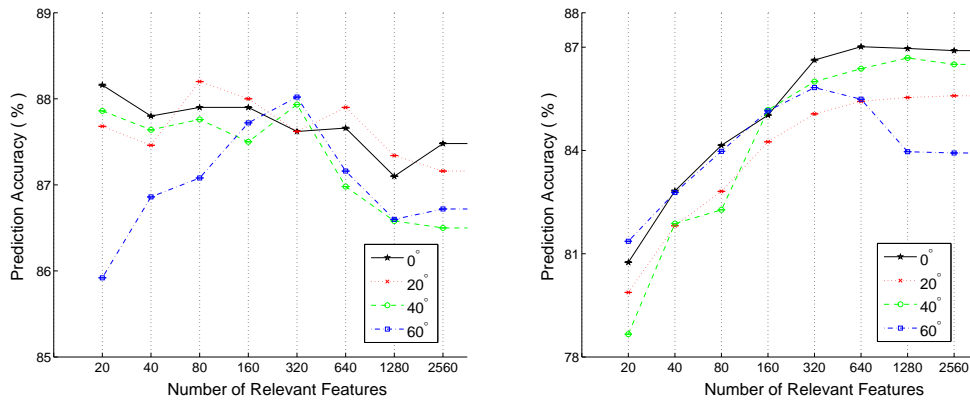


Figure 18: Mean prediction accuracies for different number of relevant features in environments with boxes in (a) and with all simple objects in (b).