Learning grasping affordances from local visual descriptors

Luis Montesano   Manuel Lopes

Abstract—In this paper we study the learning of affordances through self-experimentation. We study the learning of local visual descriptors that anticipate the success of a given action executed upon an object. Consider, for instance, the case of grasping. Although graspable is a property of the whole object, the grasp action will only succeed if applied in the right part of the object. We propose an algorithm to learn local visual descriptors of good grasping points based on a set of trials performed by the robot. The method estimates the probability of a successful action (grasp) based on simple local features. Experimental results on a humanoid robot illustrate how our method is able to learn descriptors of good grasping points and to generalize to novel objects based on prior experience.
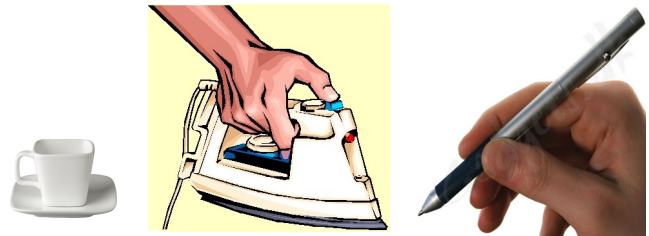

Fig. 1. Examples of graspable objects. Each object has a different point, or set of points, that allow grasping the object, e.g. the handle of the steam iron or the coffee cup. Learning grasping affordances involves discovering visual descriptors that allow to recognize and perform successfull grasps using experience even in unknown objects.

I. INTRODUCTION

Manipulation skills result from the combination of different motor and perception processes. Developing such skills is an intricate process that happens at many different levels, from sensory-motor coordination to more abstract reasoning processes. Neuroscience studies have provided insights to develop computational models on infant grasp learning [1] and generally on how humans deal with grasping [2]. In particular, there is evidence suggesting that humans easily determine, from their prior experience, what is the better way of grasping and the appropriate grasping point - i.e. the relative location of the hand and the object and the way to close the fingers around it - and can perform a grasp without any visual control, but using only visual planning [3].

In this context, affordances [4] have recently received a lot of attention from cognitive robotics as a way to encode information about the object behavior under robot actions. This type of information has been fruitfully used at different levels of abstraction including affordance based control schemes [5], tool usage [6] and imitation learning [7].

A key point of this type of knowledge is that it is based on the robot experience. This is important since it guarantees that the robot morphology is implicitly embedded in the learning process and, therefore, the resulting models depend on it. Examples of this dependency are abundant in the affordance literature in ecological psychology. Studies show how the perception of affordances varies between subjects when asked

about what is a stair [8], if a particular hole in the ground is traversable or about heaviness perception [9] . This also clearly applies to grasping tasks where the object size and weight will change human perceptions. Also, in the case of a robot the hand morphology is a crucial aspect of whether an object can be grasped or not.

This paper address the problem of learning visual descriptors for affordance properties such as graspable (see Figure 1). Learning is done through experimentation with the objects. Local visual descriptors of graspable objects are learned based on the return signal (successfull or failure) of the grasping action. Our method combines a Bayesian model of the probability of the action success based on Binomial-Beta distributions with a non-parametric kernel based approach. From a set of examples, it provides an estimate of the successfull completion of an action upon the different parts of an object. It also provides information about the confidence of this estimate. In other words, the agent is also aware of how certain his knowledge about this object is which allows to implement active approaches such as intrinsic reward [10].

We illustrate the performance of our method using Balthazar [11], a humanoid torso with an anthropomorphic hand. We recorded a set of of trials where the robot tried to grasp different objects at several points. Based on this training set, the robot successfully predicted grasping points for new objects based on prior experience.

The remainder of this paper is organized as follows. After discussing related work in the next section, section III presents our approach to grasp and grasping point learning algorithm. In Section IV, we present the experimental setup and a set of experiments that illustrates the performance of the method. Finally, in section V we discuss draw our conclusions and comment on future developments.

II. RELATED WORK

Affordance models have been widely studied in psychology and provide evidence about the existence of this knowledge in humans [8], [9]. However, there is still some ongoing discussion on the specific characteristics of this knowledge (see [12] for a review). Unfortunately, there is little evidence on how humans learn this type of knowledge.

Computational models of affordances link actions and objects, mainly in the context of manipulation tasks. They have been used to model infant grasp learning [13] where they provide context information about the possibility of executing a specific action. They have also been used to selectively focus attention during obstacle avoidance e.g. [14] or, in a Bayesian framework, as prior distributions for social interaction e.g. [7].

Discovering affordances is, inherently, a learning through experience task. Several authors have used maps to predefined input features to the desired output of an specific task. In [15], geometric features such as orientation are mapped into the motion direction resulting after poking. Self organized maps were used in [16] to associate object invariants descriptors to the successful or failure of an action. The work in [7] proposed a general model to learn multiple affordances using Bayesian networks. Other approaches use low level features such as SIFT descriptors [17] and learn tables with invariant configurations.

Since our experimental results focus on grasping, it is worth to provide some references to identify similar approaches done in the robotics literature. Most of early work used geometry to design stable grasps [18]. After this initial research the focus shifted to the relation between perception and action and several approaches tried to extract visual information to plan grasping actions: based on 2D information [19], [20], approximating to known geometric shapes [21] or using range information [22]. Other works considered learning approaches to control the manipulator to the desired grasping position based on visual servoing techniques [23] or imitation [24]. Some approaches tried to learn a mixture of perception and action. For instance, [25] learns to grasp simplified superquadric shape approximations of objects. The work in [26] learns an object classifier based on the $2D$ appearance and the camera-robot calibration using Gaussian basis functions.

Related to our approach, two previous works have learned visual descriptors for grasping called grasping points. [27] defines a reward function based on simple geometrical features of the object and some heuristics. Based on it, reinforcement learning techniques were used to learn a grasp controller with a PUMA robot and simple planar objects. The *graspable* affordance of the object at a particular point is in the Q-function. The work in [28] proposed to learn correct grasping points for a precision grip (i.e. a gripper) directly from visual features using supervised learning techniques. A good 3D grasping point is selected by fusing several 2D points and 3D planning algorithms used to reac it.
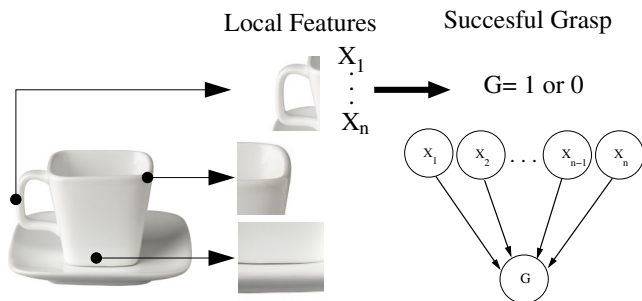

Fig. 2. Point features define the graspable affordance of an object at a particular point. Each point has different features that are mapped to the successfull grasp probability. The network shows the dependency for a particular point. Our algorithm uses kernels to locally extrapolate information from previous trials.

III. LEARNING VISUAL DESCRIPTORS FOR AFFORDANCES

This section describes our approach to learn visual descriptors for affordances. We adopt a learning through experience framework where the robot tries the action, e.g. a grasp on an object, and is able to distinguish between failures and successes. The robot perceives the object using vision and extracts a (large) set of visual descriptors. Our objective is two-fold. First, we want to learn a function to predict the probability of success of a particular action on a specific object point. Second, the learning process should also produce confidence based measures of this knowledge. This confidence is important to guide exploration for grasping points in new objects.

As a way to clarify our purpose, let us consider the case of grasping (see Fig. 1). A graspable object cannot be grasped in any way. Indeed, a particular type of grasp, e.g. a power grasp, will only succeed in some parts of the objects. Our purpose is to identify these particular points that are suitable to perform a grasp. Figure 2 illustrates the process for the coffee cup. As mentioned above, each potential grasping point is described by a set of features. Our objective is to estimate the function that maps these features to the probability of succesfully grasping the object.

In the remainder of this section, we describe our Bayesian approach to learn visual descriptors from low level features.

A. Binomial-Beta model

For the time being, let us consider a single grasping point described by the feature vector $\mathbf{x_i}$. The robot tries m times to grasp the object at this location and obtains a set of positive and negative results, S_i and U_i respectively. Each of this trials is a Bernoulli experiment. Let p_i be the probability of a successful grasp. Then, $p(S_i \mid p_i, m)$ is given by a Binomial distribution of parameters p_i and m. Our objective is to estimate the posterior distribution of p_i given the examples, $p(p_i \mid \mathbf{x}_i, S_i, U_i)$[1].

The standard Bayesian approach to compute the posterior uses the Beta distribution,

---

[1]Note that we are not estimating the maximum likelihood (or maximum a posteriori) grasping probability $\hat{p}_i$, but its full distribution.
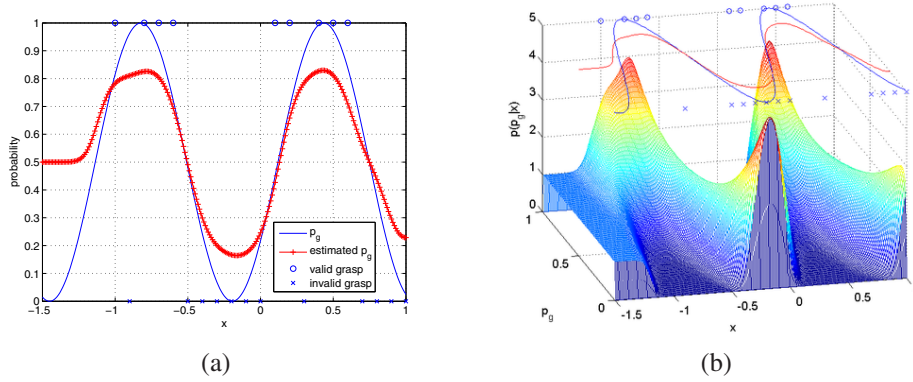
Fig. 3. Approximating a sinus varying p in a one dimensional input space. (a) Estimated mean. The 0-1 blue points are the observations generated from a Bernoulli using the true p (blue line). Failures are represented by crosses and successes by circles. The red line with marks is the approximated mean computed from the posterior. (b) Predicted posterior beta distributions for each point along x.

$$\text{Be}\,(p; \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \; p \in [0, 1] \tag{1}$$

as a prior, where $B(\alpha, \beta)$ is the Beta function. α and β are shape parameters that can be interpreted as the number of success and failed trials encoding our prior knowledge The Binomial and Beta distributions form a conjugate pair [29] and, therefore, the posterior is also a Beta distribution with updated parameters $\alpha + S_i$ and $\beta + U_i$.

B. Non-parametric smoothed beta distributions

We are now ready to introduce our algorithm to predict the grasp probability using samples taken at different places of the input space $\mathcal{X}$. We assume that the input space $\mathcal{X} \subseteq \mathbb{R}^d$ and denote input vectors as $\mathbf{x}$. The observations at a particular input vector $\mathbf{x}_i$ are the number of successful and total trials[2], $\mathbf{y}_i = (S_i, N_i) \in \mathbb{N}_0^2$. We denote a set of input vectors as $\mathbf{X}_n = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ and the corresponding observations as $\mathbf{Y}_n = \{\mathbf{y}_1, ..., \mathbf{y}_n\}$. The objective is to estimate the posterior $p(p_* \mid \mathbf{x}_*, \mathbf{X}_n, \mathbf{Y}_n)$ at an arbitrary point $\mathbf{x}_*$. Formally, we are looking for a map $f : \mathcal{X} \rightarrow \mathcal{B}_{\alpha\beta}$ where $\mathcal{B}_{\alpha\beta}$ is the space of Beta distributions parametrized by α and β.

To predict the posterior probability at point $\mathbf{x}_*$, we apply the Bayes rule,

$$\begin{aligned} p(p_* &\mid \mathbf{x}_*, \mathbf{X}_n, \mathbf{Y}_n) \\ &\propto \; p(\mathbf{Y}_n \mid p_*, \mathbf{x}_*, \mathbf{X}_n)p(p_* \mid \mathbf{x}_*, \mathbf{X}_n) \\ &= \; \prod_{i=0}^{n} p(\mathbf{y}_i \mid p_*, \mathbf{x}_*, \mathbf{x}_i)p(p_*). \end{aligned} \tag{2}$$

The last expression assumes that, given the parameter p_* at $\mathbf{x}_*$, the observed grasp results are independent of each other. The prior distribution $p(p_* \mid \mathbf{x}_*, \mathbf{X}_n)$ is just a Beta with parameters α_0 and β_0 independent of the training data. For

[2]Alternatively, we could just use the number of successful and non-successful trials.

instance, if we set both parameters to one, the prior becomes a uniform distribution.

The likelihood term $p(\mathbf{y}_i \mid p_*, \mathbf{x}_*, \mathbf{x}_i)$ represents the likelihood of the observations $\mathbf{y}_i$ given the input features $\mathbf{x}_i$ and $\mathbf{x}_*$ and the success rate p_*. To model this dependency, we have to add some structure to the model, in particular, we assume that the grasping probability is smooth given the feature space. Based on this extra structure, we take a kernel based non parametric approach [30] to extrapolate from the training dataset to unseen features. We define a kernel based diffusion for the parameters of the beta distribution between two input points $\mathbf{x}_1$ and $\mathbf{x}_2$

$$\alpha_{x_2} = K(\mathbf{x}_1, \mathbf{x}_2)\alpha_{x_1}, \qquad \beta_{x_2} = K(\mathbf{x}_1, \mathbf{x}_2)\beta_{x_1} \tag{3}$$

where $K : \mathcal{X} \times \mathcal{X} \Rightarrow [0, 1]$. Intuitively, when the kernel is a decreasing function of the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$, the diffusion process keeps the same beta parameters for the same point and decreases as the point is further in the feature space. As a result, when all the kernels vanish the prediction converges to the prior $\text{Be}(\alpha_0, \beta_0)$. The likelihood function of each individual observation is modeled as

$$p(\mathbf{y}_i \mid p_*, \mathbf{x}_*, \mathbf{x}_i) = \text{Bin}(S_{*i}; p_*, S_{*i} + U_{*i}) \tag{4}$$

where the variables $S_{*i} = K(\mathbf{x}_*, \mathbf{x}_i)S_i$ and $U_{*i} = K(\mathbf{x}_*, \mathbf{x}_i)U_i$ are the number of successful and failed grasps at point $\mathbf{x}_*$ given the grasp results observed at $\mathbf{x}_i$.

Finally, the predicted posterior probability at point $\mathbf{x}_*$ can be computed as

$$\begin{aligned} p(p_* &\mid \mathbf{x}_*, \mathbf{X}_n, \mathbf{Y}_n) \\ &\propto \; \prod_{i=0}^{n} \text{Bin}(S_{*i}; p_*, S_{*i} + U_{*i})\text{Be}(p_*; \alpha_0, \beta_0) \\ &= \; \text{Be}\left(p_*; \sum_{i=1}^{n} S_{*i} + \alpha_0, \sum_{i=1}^{n} U_{*i} + \beta_0\right) \end{aligned} \tag{5} \tag{6}$$

Fig. 4. Objects used to acquire the training dataset.

and the predicted mean grasp probability

$$\bar{p}_* = \frac{\sum_{i=1}^{n} S_{*i} + \alpha_0}{\sum_{i=1}^{n} S_{*i} + \alpha_0 + \sum_{i=1}^{n} U_{*i} + \beta_0}.$$

Figure 3 shows an example of the algorithm predictions for a one dimensional feature space. The p_i is a sinusoidal function of the input x. The 1-0 observations are generated through a binomial distribution based on the true p_i at that point. Figure 3(a) shows the true function and the predicted mean. An important feature of our algorithm is that it estimates the full posterior distribution at each feature (see Figure 3(b)). This is important because it provides an associated notion of confidence, or uncertainty, about the predicted grasping probability. The figure also illustrates how the algorithm tends to the prior distribution in the absence of measurements. In the figure, this occurs for $x < -1$ where the predicted mean value of 0.5 is due to the uniform prior.

C. Kernel parameter selection

Besides the prior Beta distribution encoding our prior knowledge about the problem, the algorithm requires to set the parameters θ that define the kernel $K(\cdot, \cdot)$. We estimate this parameters from the training data using a cross-validation approach. For each point, we compute the posterior distribution $p(p_i \mid \mathbf{x}_i, \mathbf{X}_{n,-i}, \mathbf{Y}_{n,-i})$, where $\mathbf{X}_{n,-i}, \mathbf{Y}_{n,-i}$ represent the input features and observations of the dataset except $\mathbf{x}_i$ and $\mathbf{y}_i$ respectively. Then, we use a minimum least square criterion between the predicted mean $\hat{p}$ at each point and the observed empirical one[3],

$$L_\theta(X_n, Y_n) = \sum_{i=1}^{n} (\hat{p}_i - \bar{p}_i)^2. \qquad (7)$$

In our case, we used a diagonal Gaussian kernel with independent bandwidths for each dimension of the features. The function was minimized using the *active set* method [31] and positive constraints for the kernel bandwidths. In any case, other constrained optimization methods such as the *interior point* provide similar results.

IV. EXPERIMENTAL RESULTS

This section describes the experiments performed to evaluate our method and the corresponding results. First, the experimental setup, the visual features and the robot are described. Then, we show the grasping point learning results and how to use them to grasp novel objects.

[3]We also minimized the pseudo-likelihood cross validation function, but the results where worse than using the least square one.

Fig. 5. Test objects.

A. Robot

We used Baltazar, a humanoid torso composed by a binocular head and an arm-hand [11]. The arm and the anthropomorphic robot hand have 6 and 4 degrees of freedom, respectively. For our experiments, we implemented a grasp action to pick objects from a table. It consists of two different steps. First, the robot reaches a selected grasping point using a visual servoing technique. The object is always approached from the top until until contact with the object is made. Then, the hand performs a pre-programed closure phase.

B. Computer Vision features

Recall that each potential grasping point is described by a set of features. The features need to capture enough, object, information so as to allow a smooth estimation of p. In our case, the selected features are generic filters applied directly to the saturation channel of the image. In this way, features are general, do not include any a priori object knowledge and are robust to different illumination conditions. In addition to this, they are robust, fast and easy to extract. We use the bank of 151 filters shown in Table I and Fig. 6.

TABLE I
FEATURES

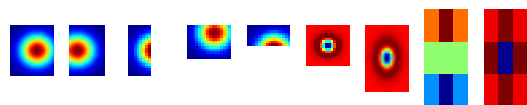| type	parameter	number
Gaussian	iso. diag. cov.	8 scales
Gaussian top	diag. cov.	8 scales
Gaussian bottom	diag. cov.	8 scales
Gaussian left	diag. cov.	8 scales
Gaussian right	diag. cov.	8 scales
Gaussian Laplacian	diag. cov.	8 scales
Gaussian	cov. $diag([d\,sd])$	8 scales$\times$6 skew
Gaussian Laplacian	cov. $diag([d\,sd])$	8 scales$\times$6 skew
Sobel	-	5 orientations
Laplacian	-	2 scales

Fig. 6. Filters used

C. Results

To evaluate the algorithm, we recorded a dataset using the 6 objects of Fig.4. The robot tried to grasp each object at several different points. The total number of actions was 550. For each of these trials we recorded the selected point, the actual point where the grasp was performed, the result of the grasp (success or failure) and the visual features of the grasped point.

First, we estimated the optimum kernel parameters for this dataset. We initialized the kernel bandwiths with random values. The number of iterations until convergence was around one hundred. Very similar results were obtained from different initializations. With slight variations, the number of kernel bandwiths different from zero was in average 12. Thus, the optimization selected a small set of the filters from the initial 151. The set of filters was stable with small variations in the weights computed for each feature.

It is important to note that there is no ground truth for our dataset. A realistic simulation is also difficult due to the hand morphology and the interactions with the objects. We present two types of evaluation. First, we computed the predicted grasp probability for each pixel of every image in the training dataset. Note that each image provided only a small set of labeled feature points to the training dataset. Since we tried the same object several times, the same object is present many times in different images. The idea is to compare the results among images from the same object and check if the method is able to generalize to the unlabeled feature points across images.

Figure 7 shows the predicted grasp probability for some images of the training dataset. The images show that predictions between different images of the same object are consistent, i.e. are similar for the same object parts in different images. Moreover, the approximation is smooth and consistent despite the fact that different images have slightly different perspectives of the objects and illumination conditions vary. The figure also shows the variance of the posterior distribution of p as a confidence measure of the prediction.

The second way to evaluate our method is to try new objects and let the robot decide where to grasp. Figure 5 shows the objects used for testing. The objects are rather different from the training ones in terms of texture, color, shape and size.

The robot tried to grasp at the point with the highest grasping probability. We repeated each experience five times for each object in Fig. 5. Table II summarizes the results. For each object, it shows the average of the grasp probability of the selected point (which is different at each trial) and the percentage of successful grasps. The comparison of these two values is not a good indicator for the performance of the algorithm. This is because the posterior is computed based on the number of trials predicted by the kernels which may be very different from the number of trials of the empirical p. Moreover, the method computes the posterior, which may be correct even if the observed p differs from the predicted one. Nevertheless, a p close to 0.5 is usually an indicator of an
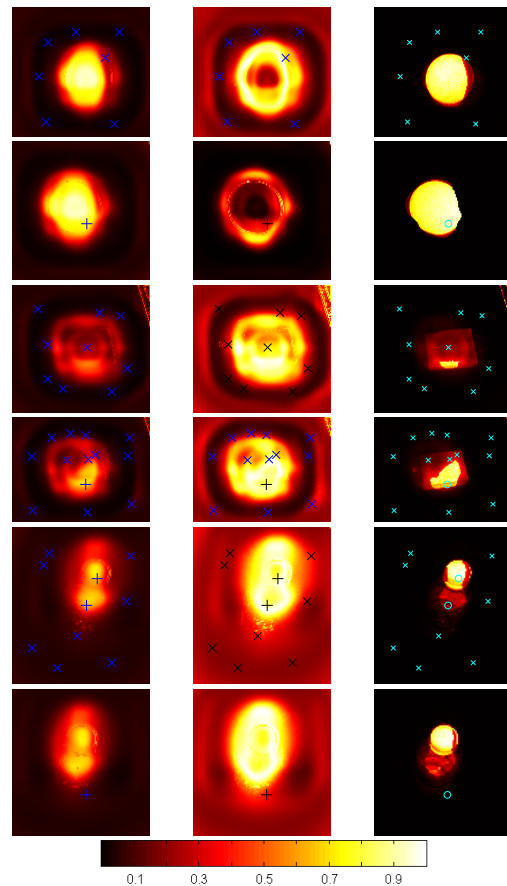

Fig. 7. Results for the big blue ball (rows 1 and 2), the pink box (rows 2 and 3) and the plastic bottle (rows 5 and 6) of the training set. Each row shows: (left) predicted grasping probability according to the color bar, (center) variance of parameter p, (right) pixels $p > 0.5$ overimposed to the saturation channel of the object image. The crosses (x) indicate failed grasps while the plus sings (+) represent successful ones.

TABLE II
RESULTS FOR 5 TRIALS ON TEST OBJECTS

Object	Grasping success(%)	
	Predicted	Observed
Ball	91%	80%
Pencil box	84%	80%
Plastic Cup	80%	80%
Tea box 1	87%	100%
Tea box 2	86%	100%
Bat	82%	80%
Salt shaker	82%	0%
Can	64%	0%
Basket	50%	80%

object whose features are far from any point in the database.

Taking into account the previous remarks, we analyze the results of Table II. The predicted point for the first five objects was always located in a similar part of the object and was a good location to perform the grasp (see Fig. 8(a) for the ball). The same happened with the bat which is a much bigger object, but whose grasping point was correctly predicted where the hand best fits it (see Fig. 8(b)).

The prediction for the next two objects was worse, since the robot could not grasp the objects at the selected points.
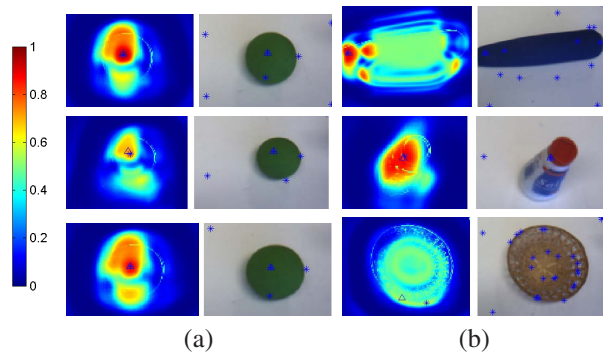
Fig. 8. Prediction for some of the test objects. The triangle and the asterisk indicate respectively the maximum predicted p and the best grasping point taken into account the control error. (a) Prediction for the ball at three different images. (b) Predictions for the bat, the salt shaker and the basket.

For the salt shaker (Fig. 8b), the variance of the posterior was low which indicates it is close to some objects in the training dataset. However, the features were not able to reflect two important differences that affected the grasp: the rigidity of the material and the conic shape which make the object slide from the hand. This was true also for the can, but here the posterior was much more uncertain reflecting the fact that the object is not very well represented in the database. Finally, the basket was marked as an almost unknown object with the posterior of the best grasping point being basically the prior (see Fig. 8b). However, the method was able to reject many negative points based on prior experience which allow the robot to grasp in a point which, in this particular case, allowed to grasp the basket.

V. CONCLUSIONS

We have presented a new algorithm to learn grasping affordances from experience. The method maps local features to the probability of success of a specific action. This map includes implicitly information about the morphology of the robot hand and represent the object *graspable* affordance. The proposed method was able to select among a set of general low level descriptors those who were relevant for the particular task. Furthermore, this is done automatically by adjusting the kernel bandwiths (i.e. the importance of the different features) based on the robot experience.

REFERENCES

[1] Erhan Oztop, Nina S. Bradley, and Michael A. Arbib. Infant grasp learning: a computational model. *Exp Brain Res*, 158:480–503, 2004.
[2] Erhan Oztop, Mitsuo Kawato, and Michael Arbib. Mirror neurons and imitation: A computationally guided review. *Neural Networks*, 19(3):254–71, 2006.
[3] Michael E. McCarty, Rachel K. Clifton, Daniel H. Ashmead, Philip Lee, and Nathalie Goubet. How infants use vision for grasping objects. *Child Development*, 72(4):973–987, 2001.
[4] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston, 1979.
[5] E. Sahin, M. Cakmak, M.R. Dogar, E. Ugur, and G. Ucoluk. To afford or not to afford: A new formalization of affordances towards affordance-based robot control. *Adaptive Behavior*, 15(5):447–472, 2007.
[6] A. Stoytchev. Behavior-grounded representation of tool affordances. In *International Conference on Robotics and Automation*, Barcelona, Spain, 2005.

[7] Luis Montesano, Manuel Lopes, Alexandre Bernardino, and José Santos-Victor. Learning object affordances: From sensory–motor coordination to imitation. *IEEE Transactions on Robotics*, 24(1):15–26, Feb. 2008.
[8] J. Konczak, H.J. Meeuwsen, and M.E. Cress. Changing affordances in stair climbing: The perception of maximum climbability in young and older adults. *Journal of Experimental Psychology: Human Perception & Performance*, 19:691–7, 1992.
[9] M.T. Turvey, K. Shockley, and C. Carello. Affordance, proper function, and the physical basis of perceived heaviness. *Cognition*, 73, 1999.
[10] Pierre-Yves Oudeyer, Frdric Kaplan, and Verana V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286, 2007.
[11] Manuel Lopes, Ricardo Beira, Miguel Praça, and José Santos-Victor. An anthropomorphic robot torso for imitation: design and experiments. In *International Conference on Intelligent Robots and Systems*, Sendai, Japan, 2004.
[12] A. Chemero. An outline of a theory of affordances. *Ecological Psychology*, 15(2):181–195, 2003.
[13] E. Oztop, N.S. Bradley, and M. Arlib. Infant grasp learning: a computational model. *Experimental Brain Research*, 158:480–503, 2004.
[14] A.C. Slocum, D.C. Downey, and R.D. Beer. Further experiments in the evolution of minimally cognitive behavior: from perceiving affordances to selective attention. In *Conference on simulation of adaptive behavior*, Paris, France, 2000.
[15] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini. Learning about objects through action: Initial steps towards artificial cognition. In *IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, 2003.
[16] I. Cos-Aguilera, L. Cañamero, and G.M. Hayes. Using a sofm to learn object affordances. In *Workshop of Physical Agents (WAF)*, Girona, Spain, 2004.
[17] G. Fritz, L. Paletta, R. Breithaupt, E. Rome, and G. Dorffner. Learning predictive features in affordance based robotic perception systems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2006.
[18] A. Bicchi and V. Kumar. Robotic grasping and contact: a review. *Proc. IEEE International Conference on Robotics and Automation*, 1:348–353, 2000.
[19] A. Morales, P. J. Sanz, and A. P. del Pobil. Vision-based computation of three-finger grasps on unknown planar objects. In *IEEE/RSJ Intelligent Robots and Systems Conference*, 2002.
[20] Jia Yan-Bin. On computing optimal planar grasps. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3:427–434, Pittsburgh, PA, USA, 1995.
[21] A.T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen. Automatic grasp planning using shape primitives. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1824 – 1829, 2003.
[22] Geoffrey Taylor and Lindsay Kleeman. Grasping unknown objects with a humanoid robot. In *Proc. 2002 Australasian Conference on Robotics and Automation*, pages 191–196, 2002.
[23] Radu Horaud, Fadi Dornaika, and Bernard Espiau. Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, 14:525–532, 1998.
[24] Jochen Triesch, Jan Wieghardt, Eric Mal, and Christoph Yon Der Malsburg. Towards imitation learning of grasping movements by an autonomous robot. In *In Proc. 3rd Gesture Workshop (GW'97)*, Lecture Notes in Artificial Intelligence, pages 73–84. Springer, 1999.
[25] Marcos Salganicoff, Lyle H. Ungar, and Ruzena Bajcsy. Active learning for vision-based robot grasping. *Machine Learning*, 23(2), 1996.
[26] Josef Pauli, Henry Hexmoor, and Maja Mataric. Learning to recognize and grasp objects. *Autonomous Robots*, 5:239–258, 1998.
[27] J. Zhang and B. Rössler. Self-valuing learning and generalization with application in visually guided grasping of complex objects. *Robotics and Autonomous Systems*, 47:117–127, 2004.
[28] Ashutosh Saxena, Justin Driemeyer, and Andrew Y. Ng. Robotic grasping of novel objects using vision. *International Journal of Robotics Research (IJRR)*, 2008.
[29] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley & Sons, Chichester, 2000.
[30] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels*. MIT, 2002.
[31] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer-Verlag, Berlin, New York, 2nd edition, 2006.