Joint Observation of Object Pose and Tactile Imprints for Online Grasp Stability Assessment

Yasemin Bekiroglu Renaud Detry Danica Kragic

***Abstract*— This paper studies the viability of concurrent object pose tracking and tactile sensing for assessing grasp stability on a physical robotic platform. We present a kernel-logistic-regression model of pose- and touch-conditional grasp success probability. Models are trained on grasp data which consist of (1) the pose of the gripper relative to the object, (2) a tactile description of the contacts between the object and the fully-closed gripper, and (3) a binary description of grasp feasibility, which indicates whether the grasp can be used to rigidly control the object. The data is collected by executing grasps demonstrated by a human on a robotic platform composed of an industrial arm, a three-finger gripper equipped with tactile sensing arrays, and a vision-based object pose tracking system. The robot is able to track the pose of an object while it is grasping it, and it can acquire grasp tactile imprints via pressure sensor arrays mounted on its gripper's fingers. We consider models defined on several subspaces of our input data – using tactile perceptions or gripper poses only. Models are optimized and evaluated with f-fold cross-validation. Our preliminary results show that stability assessments based on both tactile and pose data can provide better rates than assessments based on tactile data alone.**

I. INTRODUCTION

Object grasping and manipulation in real-world environments are, from a robotics viewpoint, uncertain processes. Despite efforts in improving autonomous grasp planners, either by learning or by building into agents sophisticated visuomotor programs, one cannot assume that a grasp will work exactly as planned. One obvious reason for this, amongst many other, is that the perceptual observations on which the planner bases its reasoning are always noisy. It is thus unlikely that the robot's fingers will come in contact with the object at the exact intended points. The object will generally move while fingers are being closed, and the final object-gripper configuration, even if geometrically similar to the intended one, may present a prohibitively different force configuration. For this reason, executing grasping actions in an open-loop system is unlikely to prove viable in real-world environments. Real-world environments will often require a closed-loop system in which perceptual feedback is constantly monitored and triggers plan corrections.

Amongst the multitude of available sensors that exist, vision and touch seem particularly relevant for grasping. Vision-driven grasping and manipulation have been extensively studied [1], [2]. Vision has typically been used to plan grasping actions, and to update action parameters as objects move. Touch-based grasp controllers have also been studied, either for controlling finger forces to avoid slippage and to prevent crushing objects [3], [4], [5], or for assessing grasp stability [6], [7].

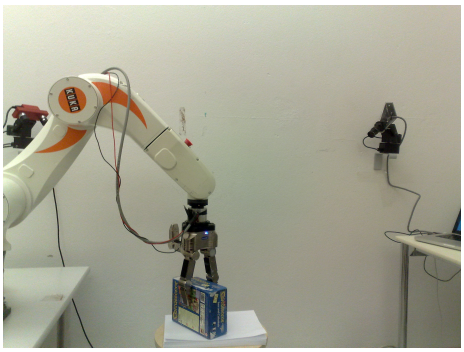

Fig. 1. Experimental robotic platform, composed of an industrial arm, a three-finger gripper equipped with tactile sensing arrays, and a camera (on the right).

In this paper, we study the joint impact of visual and tactile sensing on grasp stability assessment. Considering vision and touch separately brings valuable information on grasp stability. However, in many situations one modality can substantially help disambiguating the readings obtained from the other one. For instance, it is conceivable that for some object, two grasps approaching from different directions would yield similar tactile readings, but one would allow for robustly moving the object while the other would let the object slip away. Such situations may occur, e.g., because one of the grasps benefits from an extra gripper-object contact point in an area that is not covered by tactile sensors, or because of a different relative configuration of the grasp with respect to the center of mass of the object. Considering both modalities jointly should intuitively lead to more robust assessments. We present a platform equipped with hardware and software components which allow it to obtain 6D object pose (3D position and 3D orientation) and tactile imprint information during a grasping action, and we suggest means of using these data to *learn* a model of grasp stability.

Our robotic platform is composed of an industrial arm, a three-finger gripper equipped with tactile sensing arrays, and a vision-based object pose tracking system (see Figure 1). The robot is able to track the pose of an object while it is grasping it despite object occlusions, and it can acquire grasp tactile imprints via pressure sensor arrays mounted on its gripper's fingers.

Y. Bekiroglu, R. Detry and D. Kragic are with the Centre for Autonomous Systems, CSC, KTH, Stockholm, Sweden. yaseminb,detryr,danik@csc.kth.se.

This work was supported by EU through the project CogX, FP7-IP-027657, and GRASP, FP7-IP-215821 and Swedish Foundation for Strategic Research.

We propose to learn a stable/unstable grasp classifier from the pose and tactile data available once the robot has closed its hand around an object and is ready to attempt lifting it up. By observing the pose/touch signals issued when executing grasps demonstrated by a human, the agent learns what it feels like to grasp an object from a specific side. Once perceptual data has been acquired, an attempt to lift up the object provides the agent with a stable/unstable label for the grasp. Once a few examples are available, the agent can make predictions on the stability of a grasp before attempting to move the object. If its stability estimate is too low, the agent can for instance decide to back off and make another attempt, or possibly search locally for a more efficient grasp.

We show preliminary results on the agent's ability to gather useful data, and on its ability to learn purposeful models from these.

II. RELATED WORK

Our work is related to vision-based grasp planning, tactile sensing, and robot learning. In robotic object grasping there has been a lot of effort during the past few decades, see [8] for a recent survey. Grasp planning often utilizes grasp stability analysis that provides grasp quality measures. Analytical approaches are mostly used on grasp stability and rely on precise knowledge of the contacts between the hand and the object to estimate the stability of a grasp. Most of the grasp planning approaches tested in simulation rely on the object shape. Modelling object shape with a number of primitives such as boxes, cylinders, cones, spheres [9], or superquadrics [10] reduces the space of possible grasps. The decision about the suitable grasp is made based on grasp quality measures given contact positions. However, these kind of techniques do not provide a way of dealing with uncertainties that might arise in dynamic scenarios which can be solved using tactile feedback. To cope with the fact that the exact knowledge of the object and the hand is not available, we use tactile sensors capable of measuring a range of pressure levels. Tactile sensing has been used for various purposes in prior studies. There are recent examples which base grasp generation on visual input and use tactile sensing for closed loop control once in contact with the object. For example, the use of tactile sensors has been proposed to maximize the contact surface for removing a book from a bookshelf [11]. Application of force, visual and tactile feedback to open a sliding door has been proposed in [12]. In our work the main difference is that the tactile sensors are used to assess the stability of a grasp. Thus, rather than using the tactile data for control, we reason about grasp stability.

Learning aspects have been considered in the context of grasping mostly for the purpose of understanding human grasping strategies. In [13], it was demonstrated how a robot system can learn grasping by human demonstration using a grasp experience database. The human grasp was recognized with the help of a magnetic tracking system and mapped to the kinematics of the robot hand using a predefined lookup-table. Another approach to learning about good grasps is to use vision. However, it is impossible to measure the contact between the object and hand accurately. The system in [14] learns grasping points by using hand labeled training data in the form of image regions which indicate good grasping regions. A probabilistic decision system is then employed on previously unseen objects to determine a good grasping point or a region. In [15], the authors use vision to create probabilistic grasp affordance models for objects and refine these models through grasping. Erkan et al. [16] presented a probabilistic approach to model the success probabilities of grasp configurations obtained from visual descriptors and combined active and semisupervised learning to tackle the scarcity of labeled grasps. Current learning approaches using tactile sensors are focused on either determining the properties of objects [17], [18], [19] or object recognition [19], [20], [21], [22].

Different properties of objects give valuable information that can be further used in grasp stability analysis. In [17] the pose of the object is determined using a particle filter technique based on the tactile information gained from the contacts between a gripper and the object. Similar work was presented by Hsiao et al. [23] where object localization was performed with knowledge of tactile contacts on specific objects. [18] determine the surface type (edge, flat, cylindrical, sphere) of the tactile contact using a neural network. In [19], tactile information extracted from the sensors on a two fingered gripper is used in order to determine deformation properties of objects such as the open/closed and fill state of bottles. Also the bottle type is recognized. However, learning or analyzing such object properties through tactile sensors do not answer the question of grasp stability directly compared to the work presented here.

Work on using tactile sensors for recognition of manipulated objects has been reported rather recently. The main approach is to use multiple grasp or manipulation attempts and then learn the object through the haptic input from the manipulations or grasps. Current approaches use either one shot data from the end of the grasps [21], [22] or temporal data collected throughout the grasp or manipulation execution [19], [20]. In [21], an approach is presented to identify objects using touch sensors available in the fingertips of a gripper. The approach processes tactile images collected by grasping objects at different heights and a histogram codebook is generated using a vocabulary built by unsupervised clustering based on tactile observations. The identification is based on the histogram codebook modelling distributions over the learned vocabulary. In [22] a similar approach is taken, but with a humanoid hand. However, a more traditional approach to learning is employed by using features extracted from the tactile images which are used in conjunction with the hand joint configurations as input data for the object classifier.

To our knowledge, learning grasp controllers jointly from live visual and tactile feedback hasn't been attempted before.

III. LEARNING GRASP STABILITY

Our aim is to design an agent which can infer grasp stability estimates from the data available *before* lifting up

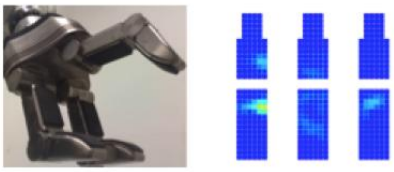

Fig. 2. We use a three-finger hand equipped with 6 tactile sensing arrays. The right side of the image shows an example of tactile readings obtained during a grasp.

an object, and to provide the agent with means of learning from experience how to make these stability assessments.

A. Kernel Logistic Regression

Formally, our agent learns an empirical representation of pose- and/or touch-conditional grasp stability probability. This model is learned from a set of examples denoted by

$$Z = \{(x_i, y_i)\}_{i=1,\ldots,n}$$

where each pair (x_i, y_i) is composed of perceptual readings $x_i \in R^d$ (pose and/or touch) and a binary stability label $y_i \in \{\text{stable, unstable}\}$. Perceptual data are read during the execution of a grasping plan, shortly after the agent closed the manipulator's fingers around the object, but before any attempt to lift or transport the object. The probability of pose- and/or touch-conditional grasp stability is modeled with kernel logistic regression as

$$P(y = \text{stable}|x; v) = \frac{1}{1 + \exp - \sum_{i=1}^{n} v_i \mathcal{K}(x, x_i)} \quad (1)$$

where $P(y = \text{stable}|x)$ is the probability of success of a grasp characterized by the tactile and/or pose vector x, $\mathcal{K}$ is a kernel function that models the similarity between two perceptual readings and v is a weight vector chosen to maximize the regularized stability probability of the data. Specifically, v is chosen as

$$\underset{v}{\operatorname{argmax}} \left\{ -\sum_{i=1}^{n} \log P(y_i|x_i; v) + c\,\text{trace}(vKv^T) \right\}, \quad (2)$$

where K is the kernel Gram matrix, with $K_{ij} = \mathcal{K}(x_i, x_j)$, and c is a constant. This problem can be solved, e.g., with Newton's method. For more details, we refer the reader to the work of Yamada et al. [24], Erkan et al. [16], and Schlkopf and Smola [25].

B. Perceptual Readings and Kernel Function

In this work, we consider perceptual signals in the form of tactile readings and/or relative object-gripper configurations. A vector x representing perceptual observations can be written as $x = (t, g)$ where t represents tactile data and g represents an object-relative gripper pose. The kernel $\mathcal{K}$ is defined as $\mathcal{K}(x_1, x_2) = \mathcal{K}_t(t_1, t_2)\mathcal{K}_g(g_1, g_2)$. Our robot platform is composed of an industrial arm and a three-finger hand. Each of the hand's finger is composed of two segments, both covered by an array of pressure sensors, yielding a total of 6 tactile arrays (see Figure 2). The tactile data is relatively high dimensional and to some extent redundant. Therefore, we start by representing the acquired data as features. Here, we borrow some ideas from image processing and consider the two-dimensional tactile patches as images. We employ image moments as a suitable representation which also reduce the dimensionality. The general parameterization of image moments for one tactile array A is given by

$$m_{p,q} = \sum_i \sum_j i^p j^q A_{ij}$$

where p and q represent the order of the moment, and i and j represent the horizontal and vertical position on the tactile patch. We compute moments up to order two, $(p + q) \in \{0, 1, 2\}$, which yields 6 numbers that model the total pressure and the distribution of the pressure in the horizontal and vertical direction. A tactile vector t, which contains moments from the six tactile pads, is composed of 6×6 numbers. The kernel function $\mathcal{K}_t$ simply corresponds to a multivariate isotropic Gaussian function

$$\mathcal{K}_t(t_1, t_2) = \mathcal{G}(t_1; t_2, \sigma_t),$$

where σ_t is a bandwidth parameter. In the next section, an optimal bandwidth is computed by cross-validation. In cases where one wishes to ignore tactile data and only take pose feedback into account, $\mathcal{K}_t(t_1, t_2)$ is simply forced to 1.

A relative object-gripper pose is computed from the pose of the hand and the pose of the object. Hand poses are simply obtained from the kinematics of the robot. Obtaining object poses is more challenging. As an object will often move while the robot is closing its hand to grasp it, the agent needs to compute the pose of the object *after* having closed the hand around it. This computation is made difficult by the partial object occlusions effected by the hand. Our aim however is not to get perfectly accurate pose information, but rather a rough idea of how the object is approached. We address this issue by tracking the movement of the object for the complete duration of the grasp. We are currently using a system which tracks the pose of a textured CAD model in a monocular video stream [26]. Tracking object textures greatly helps handling partial object occlusions and distractions induced by the hand.

An object-relative gripper pose is composed of a 3D position and 3D orientation. We define the gripper pose kernel $\mathcal{K}_g$ as the product of a position and an orientation kernel. Let us denote the decomposition of a pose g into position and orientation by p and o respectively. We define $\mathcal{K}_g$ with

$$\mathcal{K}_g(g_1, g_2) = \mathcal{G}(p_1; p_2, \sigma_p) \frac{e^{\sigma_o\, o_1^T o_2} + e^{-\sigma_o\, o_1^T o_2}}{2}$$

where $\mathcal{G}$ is a trivariate isotropic Gaussian kernel, the fraction corresponds to a pair of antipodal von-Mises Fisher distributions (Gaussian-like distribution on the rotation group [27], [28]), and the bandwidths σ_p and σ_o are fixed to allow for deviations of 20 mm and 20° respectively. For a more detailed mathematical description and motivation of $SE(3)$

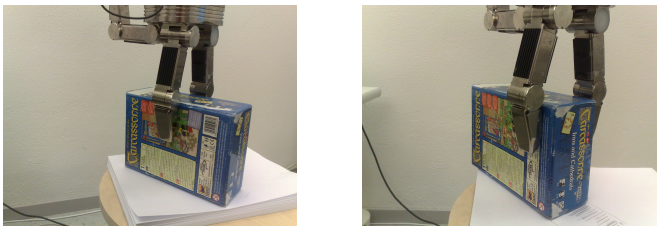

Fig. 3. Illustration of the grasps executed by the robot. The left image shows a "middle" grasp, which always succeeded, while the right image shows an "extremity" grasp which always failed.

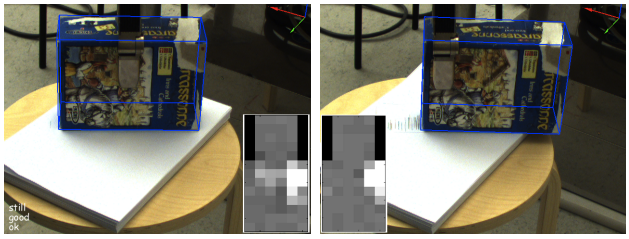

Fig. 4. Illustration of the perceptual data collected during grasps with the tracked object pose for the two grasps of Figure 3 and the tactile readings of the frontmost distal tactile pad for the two grasps of Figure 3.

kernels, we refer the reader to the work of Sudderth et al. [28]. In cases where only tactile feedback is desired, $\mathcal{K}_g(g_1, g_2)$ is set to 1.

IV. Experiment

In this section, we present perceptual data collected by the robot while grasping a box from the top, and classification rates for pose-based classification, tactile-based classification, and tactile-and-pose–based classification.

The wrist poses of the grasps executed by the robot were demonstrated by a human by teleoperation. These poses were distributed along the top side of the box (see Figure 3). As we wanted to study the stability of grasps applied along the top side of the box, only two fingers were used for grasping. The box was grasped by simultaneously closing the two fingers and continuously applying a constant closing force on all joints. A total of 50 grasps were executed, amongst which 25 were stable and 25 were unstable. Grasps applied near the middle of the top face of the box were always stable (see left image in Figure 3). As grasps were tried closer to the extremity of the box, they remained stable for a few centimeters, then abruptly became unstable. Unstable grasps were characterized by a rotation of the object when the robot tried to lift it up. Figure 4 shows the poses computed by the object tracker and the tactile reading of the two distal pads for the two grasps of Figure 3.

Stability classification was computed from the probabilistic stability model defined above (1). A grasp characterized by x was predicted to be stable if $P(y = \text{stable}|x) > \frac{1}{2}$. We computed success rates by ten-fold cross-validation. Cross-validation was run for several values of the tactile kernel bandwidth parameter σ_t (values between 0.1 and 1), and several values of the regularization constant c (see Eq. 2). Table I shows the success rates obtained in the three perceptual conditions we studied. Stability estimates computed from tactile data alone yielded a 94% rate. Estimates computed from pose data alone lead to 100% correct predictions. This result was rather surprising – we were expecting at least a few wrong prediction –, but nonetheless showed that obtaining pose information can potentially help improving tactile-based stability assessments.

	Success rate
Tactile feedback only	94%
Pose feedback only	100%
Pose and tactile feedback	100%

TABLE I

Ten-fold cross-validation of three variants of the stability classification model.

V. Discussion

In general, it is unreasonable to assume that objects will be perfectly tracked during grasps. For smaller objects, fingers will occlude a larger relative area, and pose parameters will become more noisy. By extending this experiment to other objects, we expect to find situations where neither tactile nor pose information alone are able to make a robust stability estimate, but their joint use is.

We note that in the experiment presented above, object-relative gripper poses perfectly predict stability because the object is always lying on the same face. If the object was to stand on another face, object-relative gripper poses wouldn't suffice for predicting stability. In general, the absolute orientation of the object should become part of the perceptual data to allow the model to capture the effect that gravity has on an object.

Because models rely on the pose of an object, each model that the agent learns is only usable with that particular object. It is not realistic to imagine that an agent would learn different model of every object that exists. To overcome this limitation, we project to learn models that characterize only a part of an object, and which would thus be applicable to novel objects that share the same part.

VI. Conclusion

This paper studied the viability of concurrent object pose tracking and tactile sensing for assessing grasp stability on a physical robotic platform. We presented a kernel-logistic-regression model of pose- and touch-conditional grasp success probability, and a robotic platform that can track the pose of an object while it is grasping it, and that can acquire tactile imprints of the grasps it executes. We showed that the robot is able to use data collected by human demonstrations to learn grasp stability classifiers. Our preliminary results showed that stability assessments based on both tactile and pose data can provide better rates than assessments based on tactile data alone.

REFERENCES

[1] B. Yoshimi and P. Allen, "Closed-loop visual grasping and manipulation," in *IEEE International Conference on Robotics and Automation*, 1996.

[2] D. Kragic, A. T. Miller, and P. K. Allen, "Real-time tracking meets online grasp planning," in *IEEE International Conference on Robotics and Automation*, 2001, pp. 2460–2465.

[3] A. Bicchi, J. Salisbury, and P. Dario, "Augmentation of grasp robustness using intrinsic tactile sensing," in *IEEE International Conference on Robotics and Automation*, 1989.

[4] R. Howe, N. Popp, P. Akella, I. Kao, and M. Cutkosky, "Grasping, manipulation, and control with tactile sensing," in *IEEE International Conference on Robotics and Automation*, 1990.

[5] R. Howe, "Tactile sensing and control of robotic manipulation," *Advanced Robotics*, vol. 8, no. 3, pp. 245–261, 1993.

[6] Y. Bekiroglu, D. Kragic, and V. Kyrki, "Learning grasp stability based on tactile data and HMMs," in *IEEE International Symposium in Robot and Human Interactive Communication*, 2010.

[7] Y. Bekiroglu, J. Laaksonen, J. Jorgensen, V. Kyrki, and D. Kragic, "Assessing grasp stability based on learning and haptic data," *IEEE Transactions on Robotics*, 2011, accepted to be published.

[8] B. Siciliano and O. Khatib, Eds., *Springer Handbook of Robotics, Chapter 28*. Springer, 2008, vol. Chapter 28: Grasping.

[9] 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*, 2003, pp. 1824–1829.

[10] C. Goldfeder, P. K. Allen, C. Lackner, and R. Pelossof, "Grasp Planning Via Decomposition Trees," in *IEEE International Conference on Robotics and Automation*, 2007, pp. 4679–4684.

[11] A. Morales, M. Prats, P. Sanz, and A. P. Pobil, "An experiment in the use of manipulation primitives and tactile perception for reactive grasping," in *Robotics: Science and Systems, Workshop on Robot Manipulation: Sensing and Adapting to the Real World*, Atlanta, USA, 2007.

[12] M. Prats, P. Sanz, and A. del Pobil, "Vision-tactile-force integration and robot physical interaction," in *IEEE International Conference on Robotics and Automation*, Kobe, Japan, 2009, pp. 3975—3980.

[13] S. Ekvall and D. Kragic, "Learning and Evaluation of the Approach Vector for Automatic Grasp Generation and Planning," in *IEEE Int. Conf. on Robotics and Automation*, 2007, pp. 4715–4720.

[14] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.

[15] R. Detry, E. Baseski, M. Popovic, Y. Touati, N. Krueger, O. Kroemer, J. Peters, and J. Piater, "Learning continuous grasp affordances by sensorimotor exploration," in *From Motor Learning To Interaction Learning in Robots*, 1st ed., O. Sigaud and J. Peters, Eds. Berlin, Germany: Springer-Verlag, 2010.

[16] A. Erkan, O. Kroemer, R. Detry, Y. Altun, J. Piater, and J. Peters, "Learning probabilistic discriminative models of grasp affordances under limited supervision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 1586–1591.

[17] A. Petrovskaya, O. Khatib, S. Thrun, and A. Y. Ng, "Bayesian estimation for autonomous object manipulation based on tactile sensors," in *ICRA*, 2006, pp. 707–714.

[18] A. Jiméneza, A. Soembagijob, D. Reynaertsb, H. V. Brusselb, R. Ceresa, and J. Ponsa., "Featureless classification of tactile contacts in a gripper using neural networks," *Sensors and Actuators A: Physical*, vol. 62, no. 1-3, pp. 488–491, 1997.

[19] S. Chitta, M. Piccoli, and J. Sturm, "Tactile object class and internal state recognition for mobile manipulation," in *International Conference on Robotics and Automation*, 2010.

[20] M. Schöpfer, M. Pardowitz, and H. J. Ritter, "Using entropy for dimension reduction of tactile data," in *14th International Conference on Advanced Robotics*, IEEE. Munich, Germany: IEEE, Jun 2009.

[21] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *IROS'09: Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 243–248.

[22] N. Gorges, S. E. Navarro, D. Göger, and H. Wörn, "Haptic object recognition using passive joints and haptic key features," in *In Proceedings of the IEEE International Conference on Robotics and Automation*, 2010.

[23] K. Hsiao, L. Kaelbling, and T. Lozano-Perez, "Task-driven tactile exploration," in *Proc. of Robotics: Science and Systems*, 2010.

[24] M. Yamada, M. Sugiyama, and T. Matsui, "Semi-supervised speaker identification under covariate shift," *Signal Processing*, vol. 90, no. 8, pp. 2353–2361, 2010.

[25] B. Schölkopf and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. the MIT Press, 2002.

[26] T. Mörwald, J. Prankl, A. Richtsfeld, M. Zillich, and M. Vincze, "BLORT–the blocks world robotic vision toolbox," *Best Practice in 3D Perception and Modeling for Mobile Manipulation (Workshop at ICRA 2010)*, 2010.

[27] R. A. Fisher, "Dispersion on a sphere," in *Proc. Roy. Soc. London Ser. A.*, 1953.

[28] E. B. Sudderth, "Graphical models for visual object recognition and tracking," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.