Object Identification with Tactile Sensors using Bag-of-Features

Alexander Schneider Jürgen Sturm Cyrill Stachniss
Marco Reisert Hans Burkhardt Wolfram Burgard

Abstract— In this paper, we present a novel approach for identifying objects using touch sensors installed in the finger tips of a manipulation robot. Our approach operates on low-resolution intensity images that are obtained when the robot grasps an object. We apply a bag-of-words approach for object identification. By means of unsupervised clustering on training data, our approach learns a vocabulary from tactile observations which is used to generate a histogram codebook. The histogram codebook models distributions over the vocabulary and is the core identification mechanism. As the objects are larger than the sensor, the robot typically needs multiple grasp actions at different positions to uniquely identify an object. To reduce the number of required grasp actions, we apply a decision-theoretic framework that minimizes the entropy of the probabilistic belief about the type of the object. In our experiments carried out with various industrial and household objects, we demonstrate that our approach is able to discriminate between a large set of objects. We furthermore show that using our approach, a robot is able to distinguish visually similar objects that have different elasticity properties by using only the information from the touch sensor.

I. INTRODUCTION

Touch is one of the five traditional senses that were already described by Aristotele. Humans use and rely on the sensor information from the skin while manipulating objects for a variety of sub-tasks, such as object localization, identification, and grip estimation. Additionally, there are many everyday objects that appear visually similar but can be easily distinguished using tactile sensing such as ripe versus unripe fruits. Also blind people heavily rely on their touch sense, using it to read and manipulate objects.

Accordingly, it seems very desirable to also have robots equipped with tactile sensors. Over the past years, several promising approaches have been developed on the technological or sensor side. Artificial skins that measure orthogonal pressure at comparably high spatial and temporal resolutions are typically composed of elastic, conductive, or resistive polymers, which change their electrical properties depending on the applied pressure. They can, in principle, be manufactured to cover larger parts of a robot at relatively low cost. Several research groups reported [9], [10], [11] to have successfully wrapped substantial parts of the surface of their robots using such sensors, for example, to ease human-machine interaction or to improve the robustness of object manipulation tasks.

In this paper, we show how a robotic manipulator can identify various industrial and household objects purely

All authors are with the University of Freiburg, Department of Computer Science, D-79110 Freiburg, Germany
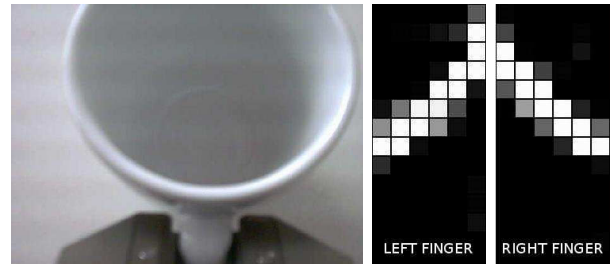{schneida, sturm, stachnis, reisert, hans.burkhardt, burgard}@informatik.uni-freiburg.de

Fig. 1: **Left:** A manipulation robot with touch-sensitive finger tips learns to distinguish a large set of objects (here: a coffee mug) solely by using its touch sense. In the image, both the cup and (at the bottom end) the two fingers of the robot's gripper are visible. **Right:** Tactile images of the sensor array in the left and the right finger. The robot is perceiving the handle of the cup.

from the observations of its touch-sensitive fingers. Given low-resolution intensity images recorded with the artificial skin, we apply k-means as unsupervised clustering on the training dataset to create a vocabulary for our bag-of-features classifier. Based on the vocabulary, we learn a codebook histogram. This histogram is a distribution over the occurrence of cluster centroids in the observed dataset. The robot is then able to use these distributions for robustly recognizing a large set of different objects requiring only a small number of grasp actions carried out at different positions. We also present an approach based on a decision-theoretic framework to minimize the number of required grasp actions. In particular, our approach efficiently estimates the expected information gain of potential future grasp actions based on the observations made during training. Experiments carried out with a large set of different objects demonstrate that our approach is able to reliably discriminate between objects. It is even possible to differentiate objects that are visually similar.

II. RELATED WORK

Tactile sensors [12], [19] are commonly defined as "a device that can measure a given property of an object or contact event through physical contact between sensor and object" that is able to sense one or more of the following modalities: pressure, normal and sheer forces, torques, slip, vibrations, or temperature. Important properties of a sensor are its spatial and temporal resolution, noise, hysteresis, creep, and aging. Different mechatronic principles have been explored in the past, such as pressure-sensitive conduc-

tive polymers [20], piezo-resistive sensors [7], piezo-electric vibration sensors [13], and capacitive sensors which can additionally measure the sheer forces [4] or temperature [3].

Tactile sensors have been used in the past to explore the 3D shape of objects [2]. Others have used tactile sensors to detect ridges and bumps in the material [14] by sliding the robotic finger over an object. Sensors based on piezo-electric vibration have been used to determine the hardness/softness of probed (biological) objects [15]. Force-sensitive fingers have been used to control the robot's position [6], i.e., to continuously keep the finger in physical contact while moving the object. It has also been shown that tactile sensors can be used to estimate the 3D pose of objects with known shapes [16]. Notably, little information is recovered from the tactile sensor in this work, resulting in multi-modal distributions due to ambiguities during the first grasps, which is a problem we are also dealing with in our work. A work relatively close to ours is that of Russel *et. al.*, who used tactile sensors for object classification [17]. Their approach extracts geometric features like point, line, or area contacts and integrates them time to classify the objects into generic classes such as boxes, spheres, cylinders etc. Later, Russel [18] showed that a similar approach can also be used for object classification using an 8-whisker tactile sensor on a robotic gripper. In contrast to their work, our method is not restricted to pre-defined geometric shapes. Rather, our method is able to recognize typical real-world objects with arbitrary shapes.

III. Touch Sensor Observations

A. Sensor principle

The robot used for gathering the data and carrying out the experiments is a RWI B21r robot equipped with a 7-DOF manipulator. The robot's end-effector is a 1-DOF gripper consisting of two fingers which both are equipped with a Weiss Robotics sensor DSA 9205 for gathering tactile images [20]. Each tactile sensor array contains 84 sensor cells arranged in 6 columns and 14 rows with a size of 24 mm by 51 mm. The maximum scanning rate for the sensor is 240 fps. Each sensor cell measures the conductivity of an elastic rubber foam above it. When a force is applied to the rubber foam, the binding polymer gets compressed thus lowering the electrical resistance of the material. The calibration of the sensor array turned out to be difficult in consequence of the sensor principle. For example, due to memory effects of the rubber foam, we took a reference measurement before the experiments were started (with no pressure on all cells). Furthermore, we normalized all measurements to the sensor's maximum response, such that we obtained for each finger a measurement $Z \in [0,1]^{6 \times 14}$.

B. Notation

In the remainder of this paper, we use the following notation for a single touch observation $\mathbf{z}$ as
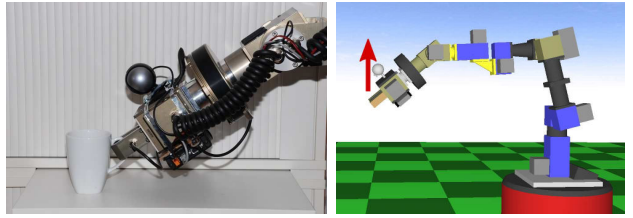
$$\mathbf{z} = < Z^{left}, Z^{right}, h, w >,$$

Fig. 2: Experimental setup. The robot grasps an object o at different positions. Each tactile observation $\mathbf{z}_i$ is then stored together with the object label in the database $\mathcal{D} = \{< \mathbf{z}_1, o_1 >, \ldots, < \mathbf{z}_N, o_N >\}$.

where $Z^{left}, Z^{right} \in [0,1]^{6 \times 14}$ are the observations of the sensor matrix of the left and right finger, while $h \in \mathbb{R}$ refers to the current height of the gripper and $w \in \mathbb{R}$ refers to the current opening width of the fingers.

C. Data Acquisition

To acquire the training data, we present a set of n different objects denoted by $O = \{1, \ldots, n\}$ to the robot, including industrial objects like metal cuboids or cylinders and house-hold objects like a cup, a toy, and a bottle. The robot grasps each object m times at different heights. This results in a set of $N = nm$ observations $\mathcal{D} = \{< \mathbf{z}_i, o_i >\}_{i=1}^N$. From this set, we sample training sets $\mathcal{D}_{\text{training}}$ for our experiments, including the true object labels, and a disjoint test sets $\mathcal{D}_{\text{test}}$ without the object labels for evaluation.

D. Distance Metric for Tactile Observations

Two images R, S (here $R, S \in [0,1]^{6 \times 14}$) can be compared by computing the Euclidean distance pixel by pixel:

$$d(R, S) = \sum_x \sum_y |r_{xy} - s_{xy}|. \tag{1}$$

To allow for small translations of the object in the robot's fingers, we do not discount vertical shifts, i.e.,

$$dist(R, S) = \min_{\tau=1,\ldots,k} (d(\mathbf{R}, shift(\mathbf{S}, \tau))). \tag{2}$$

From there, we can define a distance function for the difference between observations $\mathbf{z}_1, \mathbf{z}_2$ as

$$
\begin{aligned}
dist(\mathbf{z}_1, \mathbf{z}_2) &= \alpha \left(dist(Z_1^{\text{left}}, Z_2^{\text{left}}) + dist(Z_1^{\text{right}}, Z_2^{\text{right}}) \right) \\
&+ (1 - \alpha)|w_1 - w_2|,
\end{aligned}
\tag{3}
$$

where $\alpha \in [0,1]$ is a weighting factor determining the influence of differences in touch and finger distance. In order to circumvent scaling issues between both distance measures, we normalized both of them to have unit variance on our training dataset.

IV. The Bag-of-Features Approach

As the finger of the robot is much smaller than all of our objects, the tactile observations the robot perceives of these objects are generally only partial views. To perform the classification based on these local image patches, we apply a variant of the so-called bag-of-features approaches [21],
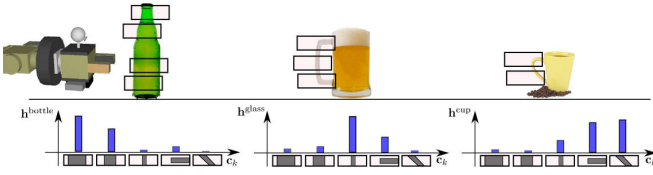
Fig. 3: Application of the bag-of-features approach with 3 objects described using 5 features. The objects are: bottle, beer glass, and coffee mug. The features in the vocabulary are thick, medium, thin vertical, thin horizontal, and thin diagonal feature. The robot grasps each object multiple times at different positions, indicated by the highlighted rectangles. This results in a characteristic histogram per object, containing the occurrence frequency of each feature in the object.

[5], [1] which have been successfully applied in the area of computer vision. Bag-of-features techniques are appealing because of both, their simplicity and power. The key idea of the bag-of-features approach is to describe the observations with a common vocabulary of features. For tactile perception, the vocabulary might include features such as "straight", "round", and "thin" observations. Given that the feature vocabulary is rich enough, the resulting feature histograms are well suited for object classification. For this purpose, a so-called codebook needs to be learned that contains the feature histograms of the trained objects. Figure 3 graphically illustrates the process of the codebook generation over objects given a vocabulary.

A. Unsupervised Creation of the Tactile Vocabulary

In practice, the appropriate vocabulary strongly depends on the objects that the robot is supposed to grasp so that predefined vocabularies will not suffice in general. Therefore, our approach is to learn a set of characteristic features automatically from the observed training data. To achieve this, our approach applies the k-means clustering algorithm directly on the observed training data $\mathbf{z} \in \mathcal{D}_{\text{training}}$. This results in k cluster centers (or centroids) $\mathbf{c}_1, \ldots, \mathbf{c}_k$, computed according to Alg. 1. The centroids are the individual words of our vocabulary. During k-means clustering, we use the distance function as defined in Eq. 3, thereby allowing the tactile images to have small vertical displacements by Eq. 2. In the remainder of this paper, we consider the set of clusters/centroids as the vocabulary that we use to describe the tactile observations.

B. Codebook Generation

As already mentioned above, the vocabulary described in the previous section is used to generate a codebook. A codebook entry $\mathbf{h}^o$ for an object o describes the distribution over centroids calculated from the training data. Each $\mathbf{h}^o$ is a histogram with k bins, $\mathbf{h}^o \in \mathbb{R}^k$. The overall set of such histograms for the codebook is denoted by H.

To build up a codebook, we initialize $\mathbf{h}^o = \mathbf{0}$ and update each bin h_i^o of $\mathbf{h}^o$ according to the observations $\mathbf{z}$ of object o

---

Initialize $c_i, i = 1, \ldots, k$ to k random $\mathbf{z}_t \in \mathcal{D}$
repeat
    forall $z_t \in \mathcal{D}$ **do**
        $b_i^t \leftarrow \begin{cases} 1 & \text{if } dist(\mathbf{z}_t, \mathbf{c}_i) = \min_j dist(\mathbf{z}_t, \mathbf{z}_j) \\ 0 & otherwise \end{cases}$
    end
    forall $c_i, i = 1, \ldots, k$ **do**
        $\mathbf{c}_i \leftarrow \sum_t b_i^t \mathbf{z}^t / \sum_t b_i^t$
    end
until $\mathbf{c}_i$ *converge* ;

Algorithm 1: The k-means clustering algorithm is used to generate a vocabulary for the bag-of-features approach.

in $\mathcal{D}_{\text{training}}$ by

$$h_i^o \leftarrow h_i^o + \exp(-dist(\mathbf{c}_i, \mathbf{z})/l), \qquad (4)$$

where l is the length scale parameter in the observation distance space. After processing all observations, the individual $\mathbf{h}^o$ must be normalized.

The key idea of the codebook is to have a compact representation of the objects that allows us to efficiently compute the likelihood that a new observation (in $\mathcal{D}_{\text{test}}$) is generated by touching a specific object o. In the next section, we explain how to compute this likelihood.

C. Observation Model

To compute the distribution over potential object classes based on an observation, we proceed as follows. By applying Bayes rule, we can write

$$p(o \mid \mathbf{z}) = \eta p(\mathbf{z} \mid o) p(o), \qquad (5)$$

where η is a normalizing constant ensuring that the left-hand side sums up to one over all o. The term $p(o)$ is the prior over the objects. In practice, this can be estimated from the training data or alternatively assumed to be uniformly distributed.

To compute the observation model $p(\mathbf{z} \mid o)$, we generate a histogram $\mathbf{h}^{\mathbf{z}}$ of a single observation $\mathbf{z}$ according to Eq. 4. As a result, we have two distributions over feature occurrences, and thus, we can express $p(\mathbf{z} \mid o)$ by computing the similarity between the feature histogram of current observation and the histogram stored in the codebook.

In the literature, there exist multiple ways for computing the similarity between histograms. Among the popular measures for comparing histograms [8] are the histogram intersection, the χ^2 distance, and the Kullback Leibler divergence (KLD). In our experiments, the histogram intersection turned out to yield the best results. This is probably due to the fact that the χ^2 distance and the KLD are heavily influenced by features with low support – an effect that can be observed frequently in our dataset. Thus, the observation model, which is based on the histogram intersection, is given by

$$p(\mathbf{z} \mid o) \propto \sum_{i=1}^{k} \min(\mathbf{h}_i^z, \mathbf{h}_i^o). \qquad (6)$$

Fig. 4: Various objects used for the experiments. **Top row:** Visual image of each object **Bottom row:** Tactile images of left and right finger of each object **From left to right:** cuboid, triangle, t-object, handle, cylinder, door key, large cup, small cup, goofy, tiger, figure, mobile, bottle, kaleidoscope, tennis ball, and soft ball.
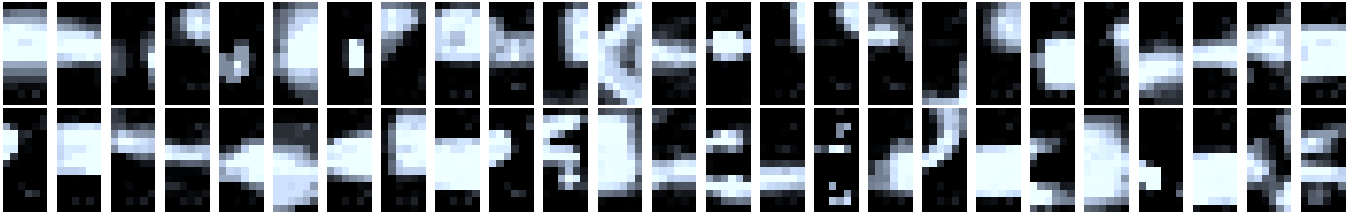

Fig. 5: Vocabulary $\mathbf{c}_1, \ldots, \mathbf{c}_k$ created using unsupervised clustering from the training data with $k = 50$ clusters. For visualization purposes, only the centroid corresponding to the tactile image of the left finger is depicted.

V. SELECTING OBSERVATION ACTIONS

To identify an object, the robot has to carry out multiple grasping actions at different height levels. Intuitively, it seems that an uninformed grasping strategy is not optimal. For example, a large number of grasps might be needed to distinguish kitchen utensils that have similar shafts. We therefore propose an informed technique based on concepts from information theory. Our approach seeks to determine the action which will provide the highest expected information gain, that is, the highest reduction of uncertainty in the posterior about potential object identity. Here, the expected information gain is the expected change of entropy in the posterior $p(o)$ of the identity o of a grasped object. The entropy is defined as

$$H(p(o)) = \int_o p(o) \log p(o) \, do. \tag{7}$$

Let $a_{1:t}$ be the actions carried out until the current time step t and let $\mathbf{z}_{1:t}$ be the corresponding observations. The robot then has to select the action a_{t+1} that provides the highest expected reduction in entropy. Let $\hat{a}$ be an action under consideration and $\hat{\mathbf{z}}$ be the corresponding observation that is obtained when carrying out $\hat{a}$.

Then, the information gain is given by

$$I(\hat{\mathbf{z}}, \hat{a}) = H(p(o \mid \mathbf{z}_{1:t})) - H(o \mid \mathbf{z}_{1:t}, \hat{\mathbf{z}})). \tag{8}$$

In general, we do not know which measurements the robot will obtain while executing action $\hat{a}$. Therefore, we have to integrate over all possible measurements $\hat{\mathbf{z}}$ to compute the expected information gain

$$E[I(\hat{a})] = \int_{\hat{\mathbf{z}}} p(\hat{\mathbf{z}} \mid \hat{a}, z_{1:t}) I(\hat{\mathbf{z}}, \hat{a}) \, d\hat{\mathbf{z}}. \tag{9}$$

Unfortunately, reasoning about all potential observations is intractable for real world applications since the number of potential measurements grows exponentially in the dimension of the measurement space. A practical approximation, however, is to sum over observations stored in the training set instead of integrating over the whole observation space:

$$E[I(\hat{a})] \approx \sum_{\hat{\mathbf{z}} \in \mathcal{D}_{\text{training}}} p(\hat{\mathbf{z}} \mid \hat{a}, z_{1:t}) I(\hat{\mathbf{z}}, \hat{a}) \, d\hat{\mathbf{z}} \tag{10}$$

Depending on the size of the training database, this sum might still be expensive to compute. To further reduce the complexity, one can easily down-sample the training set.

This approach allows us to approximate the posterior efficiently since we can directly utilize the discrete posterior about the identity of an object. The approximations substantially reduce the number of potential observations that have to be estimated by simulation compared to the number of possible measurements the sensor can generate. The ability to carry out such computations efficiently is an important prerequisite for informed action selection.

After having computed the expected information gain for each action under consideration, we select the action a_{t+1} with the highest expected utility

$$a_{t+1} = \operatorname*{argmax}_{\hat{a}} E[I(\hat{a})]. \tag{11}$$

Every time the robot has to make the decision of where to grasp next, it uses Eq. 11 to determine the action a_{t+1} with the highest expected information gain and executes it. As soon as no action provides an expected reduction of uncertainty or the robot reached a given level of certainty, the identification task is completed.

In addition to the expected reduction of the entropy, one typically has to consider a second quantity when selecting the next action. This quantity is the actual cost of carrying out an action, which needs to be traded off with the expected information gain. In our setting, however, the cost measured

k	10	20	30	40	50
% correct	58.2%	72.8%	71.7%	84.4%	83.0%

TABLE I: Parameter study on the number of clusters k.

α	0.00	0.25	0.50	0.75	1.00
% correct	66.9%	84.3%	81.0%	78.3%	76.0%

TABLE II: Parameter study on the weight α between (normalized) image distance and (normalized) gripper distance.

in terms of time needed to carry out an action can be assumed to be identical for all actions since the movements of the manipulator are carried out quickly without major differences in the time. Thus, we ignore the time needed by the manipulator for changing the height. Considering such a cost, however, can be done in a straightforward manner by adding a cost term to Eq. 11.

VI. EXPERIMENTAL RESULTS

A. Raw Data from the Touch Sensor

For testing our approach, we recorded tactile data for 16 different objects as shown in Fig. 4. The first 5 objects are industrial parts with a relatively similar shape (like a metal cube, a cylinder, and a triangle), while the latter 13 objects were household objects such as cups, toys, and bottles. We created a database of tactile observations by grasping each object on a pre-defined path (from bottom to top). Some objects were included twice in the dataset, both under $0°$, and $90°$ rotation. We obtained a set of $|\mathcal{D}| = 830$ tactile observations for 21 object labels. All experiments were then carried out on this dataset using randomized, 2-fold cross validation, resulting in two disjoint sets $\mathcal{D}_{\text{training}} \subset \mathcal{D}$ and $\mathcal{D}_{\text{test}} \subset \mathcal{D}$ of $|\mathcal{D}_{\text{training}}| = |\mathcal{D}_{\text{test}}| = 415$ samples for each run.

B. Vocabulary Creation

Before each run of our experiments, a vocabulary was created from the training data $\mathcal{D}_{\text{training}}$ by running the k-means algorithm. An example of the resulting centroids is given in Fig. 5. We tried different choices for k empirically, and found by evaluating the resulting recognition rates that $k = 50$ was a reasonable choice for the number of clusters (see Tab. I). Alternatively, one could try to find k automatically, for example, by using the Bayes information criterion (BIC). Further, we studied the influence of the weighting factor α in the distance metric of Eq. 3 (see Tab. II). For all subsequent experiments, we chose $\alpha = 0.5$ such that both the tactile images and the finger distance were considered being equally important.

C. Recognition Rates

For measuring the recognition rate, we chose an object o at random, and selected $T = 10$ random grasp observations $\mathbf{z}_{1:T}$ of that object from $\mathcal{D}_{\text{test}}$. Starting from a uniform prior $p(o)$ over all object classes, we computed the posterior $p(o \mid \mathbf{z}_{1:T})$ according to Eq. 5 and Eq. 6. From this posterior, we then selected the maximum-a-posteriori (MAP) object class

$$\hat{o} = \underset{o}{\arg\max}\, p(o \mid \mathbf{z}_{1:T}) \qquad (12)$$
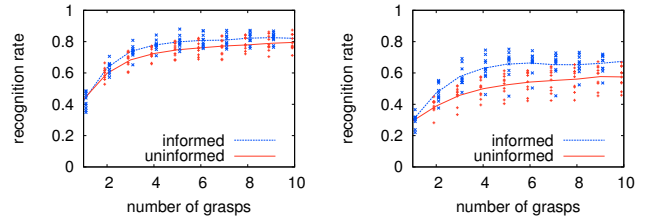

Fig. 6: Comparison of the uninformed and the informed grasping strategy depending on the number of grasp actions. **Left:** household and industrial objects (full dataset). **Right:** industrial objects only.

and compared it to the true object class o. In this way, we obtained a recognition rate of 84.6% over all 21 objects. In particular, we found that the household objects among each other were hardly ever confused (96.2%), in contrast to the industry objects (58.0%), that look very similar. The confusion matrices of these experiments are depicted in Fig. 7.

In our dataset, we also have a tennis ball and a soft ball, two objects that appear visually almost completely similar. In our experiments, we measured that these two objects could be separated from each other very well, with a recognition rate of 93.8%.

D. Active perception

We also measured whether objects can be recognized with fewer grasps when the robot selects the grasping height based on the expected information gain. We evaluated the recognition rates after each test grasp in 10 independent runs using both the uninformed and the informed grasping strategy. The results are summarized in Fig. 6. On our full dataset (household and industrial objects), using the information gain strategy performs on average 5.0% better than random grasping. In particular, one would expect that a better grasping strategy improves the recognition rates on the more difficult dataset of industrial objects. Indeed, we measured a performance gain of 18.9% of the informed strategy over the random one, raising the average recognition rate from 58.0% to 67.5%. In both experiments, a paired t-test showed significantly improved recognition rates when choosing the position that maximizes the expected information gain.

VII. CONCLUSION

In this paper, we presented a novel approach for object recognition using tactile observations obtained from the touch-sensitive fingers of a manipulation robot. Our work belongs to the class of bag-of-features techniques and maintains a probabilistic belief about the object that is currently grasped. We create a feature vocabulary for the tactile observations using k-means clustering. To recognize objects, we learn distributions over the feature vocabulary and use this to create a codebook. Furthermore, we implemented a decision-theoretic approach to active grasping that considers the expected information gain of future actions which significantly improved the recognition performance. We validated
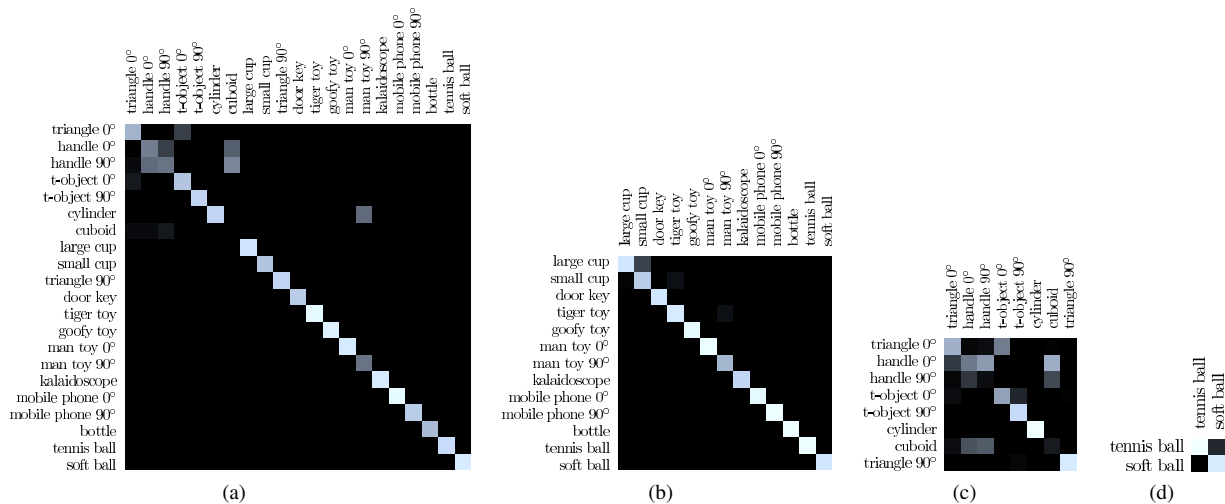
Fig. 7: Confusion matrices of object recognition after 500 object recognition trials with 10 test grasps each on different subsets of objects. (a) household & industrial objects, recognition rate: 84.6% (b) household objects only, recognition rate: 96.2% (c) industrial objects only, recognition rate: 58.0% (d) tennis ball vs. soft ball, recognition rate: 93.8%

our approach in experiments with a large range of real-world objects and obtained highly accurate recognition results.

Despite the encouraging results there are several warrants for future research. Instead of using the entire tactile images, we want to explore the possibility of using local features that are invariant to translations and orientations of the object. Additionally, we want to look at estimating the pose of the gripped object which could be beneficial during object manipulation. Also, we want to look at the time profiles of the tactile observations, for example to recognize whether and how objects are deformable.

ACKNOWLEDGMENT

REFERENCES

[1] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1475–1490, 2004.

[2] P. K. Allen. Integrating Vision and Touch for Object Recognition Tasks. *The International Journal of Robotics Research*, 7:15–33, 1988.

[3] F. Castelli. An integrated tactile-thermal robot sensor with capacitive tactile array. *IEEE Transactions on Industry Applications*, 38, 2002.

[4] C. Chuang and R. Chen. 3D capacitive tactile sensor using DRIE micromachining. In C. Cane, J.-C. Chiao, and F. Vidal Verdu, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5836, pages 719–726, July 2005.

[5] G. Csurka, L. Dance, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Proceedings of the 8th European Conference on Computer Vision, Wokshop on Statistical Learning in Computer Vision, ECCV 2004*, pages 59–74, 2004.

[6] Z. Doulgeri and S Arimoto. Force position control for a robot finger with a soft tip and kinematic uncertainties. *Robotics and Autonomous Systems*, 55(4):328 – 336, 2007.

[7] Y. Hasegawa, M. Shikida, T. Shimizu, T. Miyaji, H. Sasaki, K. Sato, and K. Itoigawa. Micromachined active tactile sensor for hardness detection. *Sensors and Actuators A: Physical*, 114(2-3):141 – 146, 2004.

[8] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3D object recognition from range images using local feature histograms. In *Proc. of the Conf. on Comp. Vision and Pattern Recognition (CVPR)*, pages 394–399, 2001.

[9] T. Hoshi and H. Shinoda. Robot skin based on touch-area-sensitive tactile element. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 3463–3468, 2006.

[10] R. Kageyama, S. Kagami, M. Inaba, and H. Inoue. Development of soft and distributed tactile sensors and the application to a humanoid robot. In *IEEE International Conference on Systems, Man, and Cybernetics 1999 (IEEE SMC 99)*, volume 2, pages 981 – 986, 1999.

[11] O. Kerpa, K. Weiß, and H. Wörn. Development of a flexible tactile sensor system for a humanoid robot. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, volume 1, 2003.

[12] M. H. Lee and H. R. Nicholls. Review article tactile sensing for mechatronics–a state of the art survey. *Mechatronics*, 9(1):1 – 31, 1999.

[13] K. Motoo, T. Fukuda, F. Arai, and T. Matsuno. Piezoelectric vibration-type tactile sensor with wide measurement range using elasticity and viscosity change. *Advanced Robotics*, 24(3), 2006.

[14] A.M. Okamura and M. R. Cutkosky. Haptik exploration of fine surface features. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1999.

[15] S. Omata, Y. Murayama, and C. E. Constantinou. Real time robotic tactile sensor system for the determination of the physical properties of biomaterials. *Sensors and Actuators*, 112(2-3):278 – 285, 2004.

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

[17] R.A. Russell. Object recognition by a 'smart' tactile sensor. In *Proceedings of the Australian Conference on Robotics and Automation*, Melbourne, Australia, 2000.

[18] R.A. Russell and J.A. Wijaya. Object location and recognition using whisker sensors. In *Proceedings of the Australian Conference on Robotics and Automation 2003*, 2003.

[19] J. Tegin and J. Wikander. Tactile sensing in intelligent robotic manipulation - a review. *Industrial Robot: An International Journal*, 32, 2005.

[20] K. Weiss and H. Wörn. The Working Principle of Resistive Tactile Sensor Cells. *IEEE International Conference on Mechatronics & Automation*, 2005.

[21] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. In *Proc. of the Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, 2006.