



HAL
open science

Zero-Shot Object Recognition Based on Haptic Attributes

Zineb Abderrahmane, Gowrishankar Ganesh, Andrea Cherubini, André Crosnier

► **To cite this version:**

Zineb Abderrahmane, Gowrishankar Ganesh, Andrea Cherubini, André Crosnier. Zero-Shot Object Recognition Based on Haptic Attributes. ICRA Workshop The Robotic Sense of Touch, May 2017, Singapour, Singapore. hal-01523030

HAL Id: hal-01523030

<https://hal.science/hal-01523030>

Submitted on 16 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Zero-Shot Object Recognition Based on Haptic Attributes

Zineb Abderrahmane¹, Gowrishankar Ganesh^{1,2}, Andrea Cherubini¹ and André Crosnier¹

Robots operating in household environments need to recognize a variety of objects. Several touch-based object recognition systems have been proposed in the last few years [2]–[5]. They map haptic data to object classes using machine learning techniques, and then use the learned mapping to recognize one of the previously encountered objects. The accuracy of these proposed methods depends on the mass of the the training samples available for each object class. On the other hand, haptic data collection is often system (robot) specific and labour intensive. One way to cope with this problem is to use a *knowledge transfer* based system, that can exploit object relationships to share learned models between objects. However, while knowledge-based systems, such as zero shot learning [6], have been regularly proposed for visual object recognition, a similar system is not available for haptic recognition.

Here we developed [1] the first haptic zero-shot learning system that enables a robot to recognize, using haptic exploration alone, objects that it encounters for the first time. Our system first uses the so called *Direct Attributes Prediction (DAP) model* [7] to train on the semantic representation of objects based on a list of haptic *attributes*, rather than the object itself. The attributes (including physical properties such as shape, texture, material) constitute an intermediate layer relating objects, and is used for knowledge transfer. Using this layering, our system can predict the attribute-based representation of a new (previously non-trained) object and use it to infer its identity.

A. System Overview

An overview of our system is given in Fig. 1. Given distinct training and test data-sets Y and Z , that are described by an attribute basis a , we first associate a binary label a_m^o to each object o with $o \in Y \cup Z$ and $m = 1 \dots M$. This results in a binary object-attribute matrix K . For a given attributes list during training, haptic data collected from Y are used to train a binary classifier for each attribute a_m . Finally, to classify a test sample x as one of Z objects, x is introduced to each one of the learned attribute classifiers and the output attributes posteriors $p(a_m | x)$ are used to predict the corresponding object, provided that the ground truth is available in K .

This extended abstract is a summary of submission [1]

¹Interactive Digital Human group IDH, Laboratory for Computer Science, Micro-electronics and Robotics LIRMM, University of Montpellier CNRS, 860 rue Saint Priest, 34090 Montpellier, France. firstname.lastname@lirmm.fr

²CNRS-AIST JRL UMI3218/RL 1-1-1 Umezono, 305-8560 Tsukuba, Japan. g.ganesh@aist.go.jp

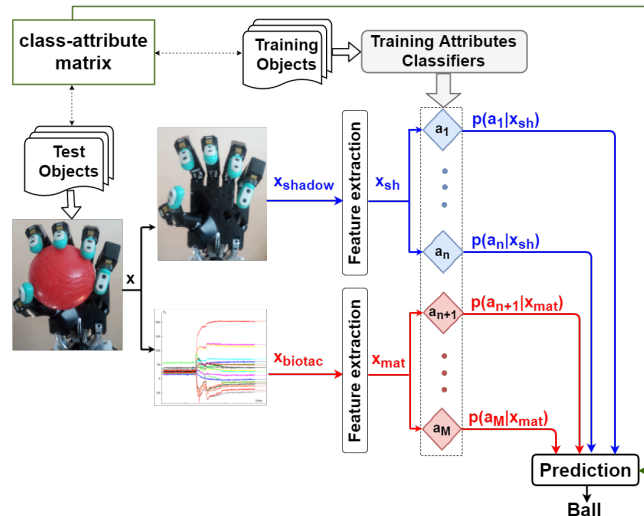


Fig. 1. Direct Attribute Prediction for Zero-Shot Haptic Recognition

B. Experimental Setup

To collect haptic data, we use the *Shadow* anthropomorphic robotic hand equipped with a *BioTac* multimodal tactile sensor on each fingertip. We developed a force-based grasp controller that enables the hand to enclose an object. The joint encoder readings provides us with information on object shape, while the *BioTac* sensors provides us with information about objects material, texture and compliance at each fingertip¹.

In order to find the appropriate list of attributes describing our object set (illustrated in Fig. 2), we used online dictionaries to collect one or multiple textual definitions of each object. From this data, we extracted 11 haptic adjectives, or descriptions that could be “felt” using our robot hand. These adjectives served as our attributes: *made of porcelain*, *made of plastic*, *made of glass*, *made of cardboard*, *made of stainless steel*, *cylindrical*, *round*, *rectangular*, *concave*, *has a handle*, *has a narrow part*. We grouped the attributes into *material attributes*, and *shape attributes*.

During the training phase, we use the *Shadow* hand joint readings x_{sh} to train an SVM classifier for each shape, and *BioTacs* readings x_b to train an SVM classifier for each material attribute. SVM training returns a distance $s_m(x)$ measure for each sample x that gives how far x lies from the discriminant hyper-plane. We transform this score to an attribute posterior $p(a_m | x)$ using a sigmoid function.

¹A video of the experiments has been attached to this paper and can be found on the IDH YouTube channel: <https://youtu.be/GaGHZXHBDG4>



Fig. 2. Disjoint training (left) and test (right) object sets.

C. Zero-Shot Learning

The last step is to classify a test sample x as one of the test objects $Z = \{z_1, \dots, z_L\}$. According to the DAP model [7], we introduce x to each attribute classifier to obtain a list of attributes posteriors $\{p(a_1 | x), \dots, p(a_M | x)\}$. From these, we infer the posterior of a test object z_l :

$$\begin{aligned} p(z_l | x) &= \sum_{\mathbf{a} \in \{0,1\}^M} p(z_l | \mathbf{a}) p(\mathbf{a} | x) \\ &= \sum_{\mathbf{a} \in \{0,1\}^M} p(\mathbf{a} | z_l) \frac{p(z_l)}{p(\mathbf{a})} p(\mathbf{a} | x). \end{aligned} \quad (1)$$

Since we have the attributes vector \mathbf{a}^{z_l} associated to z_l , $p(\mathbf{a} | z_l) = 1$ if $\mathbf{a} = \mathbf{a}^{z_l}$ and 0 otherwise. Assuming independence between attributes, we obtain $p(\mathbf{a} | x) = \prod_{m=1}^M p(a_m | x)$. Replacing in (1) yields:

$$p(z_l | x) = \frac{p(z_l)}{p(\mathbf{a}^{z_l})} \prod_{m=1}^M p(a_m^{z_l} | x). \quad (2)$$

We use MAP estimator to classify x to the best matching class $f(x) = z_l$:

$$f(x) = \underset{z_l}{\operatorname{argmax}} p(z_l | x). \quad (3)$$

D. Accounting for Inter-grasp variability in haptic data

Grasps are usually planned for a given object shape. Here we aim for the opposite; to recognize the shape (and material) from a grasp. Our system is thus constrained to work with unplanned rudimentary grasps, which in turn can lead to large variability in the observed haptic attributes. First, it is possible that the size of the object is large and that the robot grasp is not able to enclose the object completely. This can lead to the absence of some attributes in each grasp. Second, the object may be very small, or the grasp orientation improper, such that certain fingers don't touch the object, again leading to variability of haptic information across grasps. Furthermore, even a properly enclosed object may be made of different materials leading to different BioTacs readings from each finger and each grasp.

To cope with these problems, we propose an extension of the DAP model that deals with absent attributes and allows for the integration of information from different fingers. We refer to this extended model as Local DAP, or LDAP. LDAP assumes that if the system reports that an attribute is absent from the object, then the attribute posterior is replaced by

	ball	rec. cont.	tube	blender	bowl	glass	plas. cup	msr. cup	jar	salter
ball	0.6	0.1	0.2	0	0	0.1	0	0	0	0
rec. cont.	0.06	0.75	0.19	0	0	0	0	0	0	0
tube	0	0	0.9	0	0	0	0	0	0	0.1
blender	0	0	0	0.53	0	0	0.2	0	0.13	0.13
bowl	0	0	0	0	1	0	0	0	0	0
glass	0	0	0.3	0	0	0.7	0	0	0	0
plas. cup	0	0	0.3	0	0	0.1	0.5	0	0	0.1
msr. cup	0	0	0.46	0	0	0.08	0.31	0.15	0	0
jar	0.1	0.3	0.4	0	0	0	0	0	0.2	0
salter	0	0	0	0.2	0	0	0	0	0	0.8

Fig. 3. Confusion matrix of LDAP

a uniform probability to express our uncertainty about the presence of this attribute. Moreover, the posteriors we obtain for a material attribute from the B contacting fingers are averaged to obtain a global posterior of the whole grasp. We can formulate these two modifications in the following formulas, for material attributes:

$$p(a_m = 1 | x) = \frac{1}{B} \sum_{b=1}^B \max(0.5, p(a_m = 1 | x_b)). \quad (4)$$

and for shape attributes:

$$p(a_m = 1 | x) = \max(0.5, p(a_m = 1 | x_{sh})), \quad (5)$$

E. Evaluation

Fig. 3 gives the confusion matrix when applying LDAP on our test object set. Note again that these objects were distinct from the objects the robot trained on. Our approach shows good performance on most objects. It however does less well with some objects such as the measuring cup and the jar, which was partly expected due to the heterogeneity of the objects itself, and the similarity between the objects.

F. Conclusion

In this work, we designed an attribute-based zero-shot haptic recognition system for recognizing a set of household objects. We used our robotic setup, namely the Shadow hand and BioTac sensors, to collect haptic data from each object, and the DAP model to classify a test sample to one of a set of classes, that the robot has never been trained on.

We are now working to improve and extend the system to better combine information from multiple grasps, consider real-valued attributes [8], and to integrate our zero-shot learning architecture with previous state of the art methods for recognizing known objects.

REFERENCES

- [1] Z. Abderrahmane, G. Ganesh, A. Cherubini, and A. Crosnier, "Zero-shot object recognition based on haptic attributes," 2017, submitted to IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS) and IEEE Robotics and Automation Letters.
- [2] D. Tanaka, T. Matsubara, K. Ichien, and K. Sugimoto, "Object manifold learning with action features for active tactile object recognition," in *IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS)*, 2014, pp. 608–614.

- [3] M. V. Liarokapis, B. Calli, A. J. Spiers, and A. M. Dollar, "Unplanned, model-free, single grasp object classification with underactuated hands and force sensors," in *IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS)*, 2015, pp. 5073–5080.
- [4] M. Kaboli, R. Walker, G. Cheng, *et al.*, "In-hand object recognition via texture properties with robotic hands, artificial skin, and novel tactile descriptors," in *IEEE-RAS Int. Conf. on Humanoid Robots*, 2015, pp. 1155–1160.
- [5] H. Liu, D. Guo, and F. Sun, "Object recognition using tactile measurements: Kernel sparse coding methods," *IEEE Trans. on Instrumentation and Measurement*, vol. 65, no. 3, pp. 656–665, 2016.
- [6] M. Palatucci, D. Pomerleau, G. E. Hinton, and T. M. Mitchell, "Zero-shot learning with semantic output codes," in *Advances in Neural Information Processing Systems*, 2009, pp. 1410–1418.
- [7] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 951–958.
- [8] D. Parikh and K. Grauman, "Relative attributes," in *Computer Vision (ICCV), 2011 IEEE Int. Conf. on*, 2011, pp. 503–510.