Haptic Zero-Shot Learning: Recognition of Objects Never Touched Before

Zineb Abderrahmane, Gowrishankar Ganesh, André Crosnier, Andrea Cherubini

Haptic Zero-Shot Learning: Recognition of Objects Never Touched Before

Zineb Abderrahmane[a], Gowrishankar Ganesh[a,b], André Crosnier[a], Andrea Cherubini[a]

[a] *Interactive Digital Human IDH, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM), University of Montpellier CNRS, 860 Rue Saint Priest, 34095 Montpellier, France*
[b] *CNRS-AIST JRL UMI3218/RL 1-1-1 Umezono, 305-8560 Tsukuba, Japan*

Abstract

Object recognition is essential to enable robots to interact with their environment. Robots should be capable, on one hand of recognizing previously experienced objects, and on the other, of using the experienced objects for learning novel objects, i.e. objects for which training data are not available. Recognition of such *novel objects* can be achieved with Zero-Shot Learning (ZSL). In this work, we show the potential of ZSL for haptic recognition. First, we design a zero-shot haptic recognition algorithm and, using the extensive PHAC-2 database [1] as well as our own, we adapt, analyze and optimize the ZSL for the challenges and constraints characteristic of haptic recognition. Finally, we apply the optimized algorithm for haptic recognition of daily-life objects using an anthropomorphic robot hand. Our algorithm enables the robot to recognize eight of the ten novel objects handed to it.

Keywords: Haptic recognition, Zero-shot learning, Attribute-based description, Robotic hand.

1. Introduction

Transferring human perceptual capabilities to robots can enhance their autonomy, and hence facilitate their integration in human environment. Object recognition is a powerful human capability that, when transferred to a robot, can optimize grasping and manipulation. For instance, recognizing a mug and its functionalities can guide a robot to hold it vertically to avoid spilling the liquid.

Object recognition has been extensively studied in computer vision and tremendous advances have been made [2]. Visual recognition of an object however, requires it to be in the camera field of view, and is limited by lighting constraints and visual occlusions. One way to overcome these constraints is by the integration of visual recognition with additional sources of information, such as haptics; i.e., touch and proprioception. Haptics can not only help overcome the constraints associated with visual recognition, but also provide additional information regarding the shape, size, texture and compliance of an object, that can arguably improve recognition.

An object encountered by a robot can be one that it has *enough* experience with (that it has enough training data on). In this case, the object can be recognized using the available haptic recognition algorithms that use multi-class classifiers to efficiently deal with challenging robotic setups [3, 4] and object sets [5, 6], using various machine learning and optimization techniques [7, 8, 9]. On the other hand, an encountered object may be one with sparse, or in the worst case, no training data [10, 11, 12, 13]. In our study, we focus on the problem of recognizing new objects that have never been experienced before. To the best of our knowledge, this problem has never been addressed in regard to haptic recognition.

Recognition of novel objects that have not been trained on, is referred to as *Zero-shot learning* or ZSL. The key idea underlying ZSL is the exploitation of object relationships in order to transfer the knowledge learned from known objects, to recognize novel ones. In this work, we adopt an attributes-based approach. An attribute is a high-level object property that may represent, for example object size (e.g. small), or object texture (eg. rough). By describing all objects using attributes, the system can be trained to learn how to map haptic data to attributes. This mapping is

[1] Email adresses: firstname.lastname@lirmm.fr
[2] Email adress:g.ganesh@aist.go.jp

used to recognize novel objects solely given their attribute-based description. For example, the system can recognize a bottle given its description: *a glass or plastic container that has a narrow neck and usually has no handle*[3], if it has learned from known objects the meaning of each underlined attribute.

Different studies (e.g.[7]) have shown that vision techniques can be effectively applied to haptics. This motivated us to use the Direct Attributes Prediction (DAP) model, proposed by [14] for vision-based ZSL, on haptic data. First, we adapt DAP for the haptic constraints and develop a zero-shot learning (ZSL) algorithm for haptic recognition using the state of art PHAC-2 database. This database provides haptic data and attribute-based descriptions for 60 objects that cover a wide range of physical properties. Next, we implement the developed ZSL framework on our anthropomorphic robot hand equipped with tactile fingertip BioTac sensors. Our robot can recognize novel objects by grasping them, and without requiring vision or any prior information about the object shape or orientation.

This paper is organized as follows: section 2 reviews the state of art studies on haptic recognition and ZSL. Section 3 presents the theoretical background of ZSL and the attribute-based approach. Section 4 investigates the application of ZSL to haptic data using the PHAC-2 database. Section 5 details our proposed solutions to adapt the theoretical framework to object recognition with a robot hand. Finally, section 6 provides conclusions and future perspectives.

2. Related work

During the last three decades, there has been a growing interest in exploiting haptic data for object recognition. An early approach [15, 16, 17, 18, 19, 20] uses contact points and robot joint readings to perform shape-based recognition. Recent studies, instead, consider tactile contact patterns as tactile images [4, 6, 7, 21, 22, 23, 24] and make use of the rich state of art on pattern recognition. In this regard, studies have also focused on developing sophisticated tactile sensors and thus integrating new modalities such as texture [25], material [26] and compliance [27] for recognition. In addition, new feature extraction and classification techniques such as deep learning [6] and sparse coding [28, 9] were shown to improve recognition accuracy. Recent studies have also targeted object recognition in special environments (e.g., underwater [29]), dealing with data sparsity and noise [30], recognizing with a single-grasp [31], or incrementally [32], and exploring actively [33, 34].

An important but rarely discussed issue in haptic recognition is the high cost of training data collection, which is practically unfeasible for the plethora of daily life objects. Traditional multi-class classifiers on the other hand, require large data population for good recognition performance. When using a traditional multi-class classifier, the unbalance in the number of training data between objects, under-fits those with few data, and omits those with no data at all. On the other hand, spending time on collecting sufficient training data for all objects can obviously be extremely hardware consuming. *Knowledge transfer*, by reusing knowledge previously learned from other objects, can efficiently cope with the lack of data. In case of under-represented classes, this solution is known as one-shot learning. It has been applied to haptic recognition based on texture properties by Kaboli et al. in [35, 36] showing that models learned from ten objects can be used to recognize twelve new textures having one or few training samples. The authors utilize knowledge transfer to avoid the long data collection phase.

In the extreme case, under-represented objects may have no training data. This case is called *zero-shot learning* or ZSL. Note that a multi-class classifier will fail to classify samples to classes having zero training data. ZSL has been extensively discussed for visual recognition [37] but never so in regard to haptic object recognition. Lampert et al. proposed the first attribute-based framework in [14] for visual recognition. They built the «Animals with Attributes» image database of 50 animal classes separated into 2 disjoint sets: training and test sets. A list of 85 human defined binary attributes (e.g. white, furry, small) were used to describe all animals. A series of studies followed. In [38], multiple techniques for mining attributes and class-attribute associations using multiple linguistic knowledge bases were compared. Yu et al. [10] proposed a new framework based on generative models. The authors of [39] studied attribute learning in large-scale datasets.

All these cited works use binary attributes, [40, 13] generalized to real-valued attributes. Kankuekul et al. [41] handled attributes inconsistency when learned incrementally from different persons. To minimize human attribute definition and labeling effort, [42] proposed an automatic method to design non-semantic data-driven attributes. The

[3] Merriam-Webster dictionary

authors of [43] considered attributes unreliability and proposed a statistic solution to leverage errors in attributes. In [44], the two stages of learning attributes and inference were replaced by a simplistic implementation.

In this work, we choose to use semantic binary attributes as they are intuitive and yet shown to perform well on visual data [14]. To define a set of attributes describing the haptic sensation of objects, we refer to an important work for describing haptic sensation using words [1]. This work proposed a framework to describe 60 different objects using a list of 24 binary haptic adjectives based on human blind haptic exploration of objects. We believe this is an important step toward improving the answer of a robot while touching objects: instead of returning the name or the identity of the object, the authors propose to return a vector of binary values indicating the presence of each haptic adjective in the object. In addition to the multiple applications of this adjective-based description presented in [1], we investigate its application to zero-shot learning.

3. Zero-shot learning

3.1. Problem formulation

ZSL consists of training and testing a recognition system by using disjoint training and test class sets. Specifically, the system is trained on $D_{train} \subset X \times Y$ and tested on $D_{test} \subset X \times Z$, where X is the feature space in which data is represented and $Y = \{y_1, \ldots, y_N\}$ and $Z = \{z_1, \ldots, z_L\}$ two disjoint sets of classes[4]. A multi-class classifier will fail to learn a mapping function $f : X \longrightarrow Z$ from D_{train} in such a case since $Y \cap Z = \emptyset$.

3.2. Solution overview

The key idea behind ZSL is to exploit relationships between Y and Z to transfer knowledge learned from Y to recognize objects in Z through a layer of attributes, as illustrated in Fig. 1. All objects from $O = Y \cup Z$ are described using a set of attributes $A = \{a_1, \ldots, a_M\}$ (e.g. $A = \{thin,\ soft,\ rigid\}$). This results in the object-attribute matrix $\boldsymbol{K}$ (see Fig. 1a), in which the row i is the attributes-based description $\boldsymbol{a}^{o_i} = \left[a_1^{o_i}, \ldots, a_M^{o_i}\right]$ of an object $o_i \in O$, such that $a_m^{o_i} = 1$ if attribute a_m is present in object o_i, and $a_m^{o_i} = 0$ otherwise (e.g. vector $\boldsymbol{a} = [1, 0, 1]$ for object class pencil). Then, D_{train} and attribute-based descriptions of training objects $\{\boldsymbol{a}^{y_n}, y_n \in Y\}$ are used to learn a mapping from X to A: for any $x \in X$, this mapping predicts a probabilistic attribute-based description $p(a_1|x), \ldots, p(a_M|x)$, where $p(a_m|x)$ is the probability of the presence of a_m in x (e.g. $p(thin|x) = 0.8$, $p(soft|x) = 0.2$, $p(rigid|x) = 0.5$ for x collected from a pencil)(see Fig. 1b). Finally, for a test sample $x \in X$, the returned posterior $p(a_m|x)$ of each attribute a_m is used by a DAP model [14] along with the attribute-based descriptions of test objects $\{\boldsymbol{a}^{z_l}, z_l \in Z\}$ to infer the final classification $z_x = f(x)$ (see Fig. 1c).

In this scenario the cost of adding a new object to the test set is dramatically lower than with traditional learning. In fact, it breaks down to adding a row (the attribute-based object description) to matrix $\boldsymbol{K}$. This is a very low cost as compared to collecting haptic data for the new object to retrain the multi-class classifier.

3.3. Direct Attributes Prediction (DAP)

The attribute-based zero-shot recognition system of [14] proposed two models: Direct Attributes Prediction (DAP) and Indirect Attributes Prediction (IAP). When comparing models recognition accuracies [14, 12], the difference is relatively small, with a slightly better accuracy given by DAP. Thus, in this work, we focus only on DAP.

DAP is a probabilistic model that makes use of attributes learned from trained objects Y, to classify a test sample $x \in X$ as one of a set of novel objects Z, using object-attribute matrix $\boldsymbol{K}$. During the training phase, DAP learns a mapping from X to A. By assuming attribute independence, for each attribute a_m, an independent probabilistic binary classifier $g_m : X \longrightarrow [0, 1]$ is trained with $D_{train}^m = \{(x, a_m^y)$ s.t. $(x, y) \in D_{train}\}$. Each g_m returns the attribute posterior $p(a_m|x)$ for a given $x \in X$.

During the test phase, to recognize a sample x as one of Z objects, DAP computes a posterior probability $p(z_l \mid x)$ for each object $z_l \in Z$. For this, both the object's attribute-based description $\boldsymbol{a}^{z_l}$ (provided by matrix $\boldsymbol{K}$) and its

[4]Since we study object recognition, a class here represents an object category.

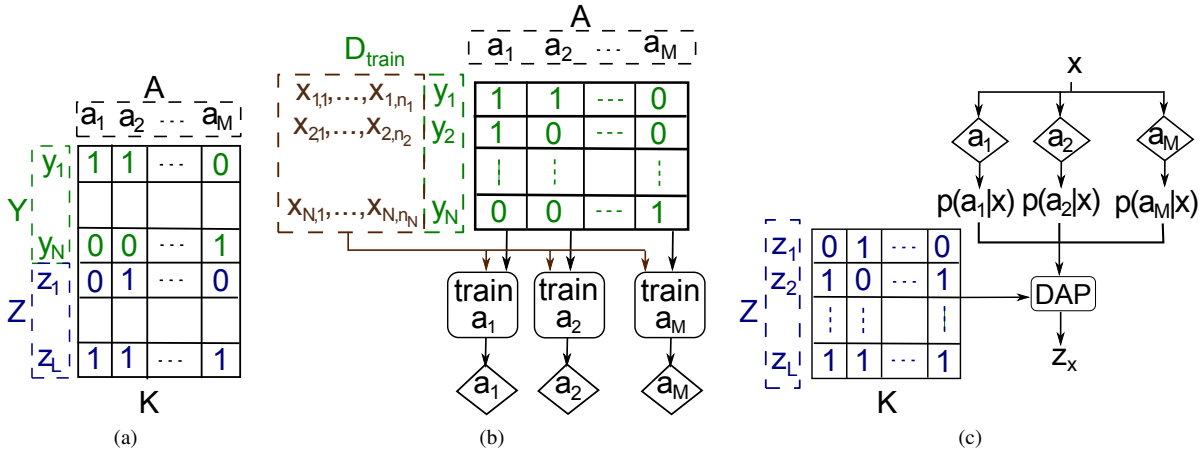


Figure 1: Attribute-based zero-shot learning: system overview.

posterior probability $p(\boldsymbol{a}^{z_l}|x)$ are needed. To compute the z_l object posterior, the attributes layer is introduced and Bayes' rule is applied to obtain:

$$\begin{aligned} p(z_l \mid x) &= \sum_{\boldsymbol{a}\in\{0,1\}^M} p(z_l \mid \boldsymbol{a})\; p(\boldsymbol{a} \mid x) \\ &= \sum_{\boldsymbol{a}\in\{0,1\}^M} p(\boldsymbol{a} \mid z_l)\frac{p(z_l)}{p(\boldsymbol{a})} p(\boldsymbol{a} \mid x), \end{aligned} \tag{1}$$

Since each object z_l is associated with a discriminating attributes vector $\boldsymbol{a}^{z_l}$, $p(\boldsymbol{a} \mid z_l) = 1$ if $\boldsymbol{a} = \boldsymbol{a}^{z_l}$ and 0 otherwise. Replacing in (1) yields:

$$p(z_l \mid x) = \frac{p(z_l)}{p(\boldsymbol{a}^{z_l})} p(\boldsymbol{a}^{z_l} \mid x), \tag{2}$$

where test class priors are considered uniform. DAP assumes attributes independence so $p(\boldsymbol{a}^{z_l}|x) = \prod_{m=1}^{M} p(a_m^{z_l} \mid x)$ and (2) becomes:

$$p(z_l \mid x) = \frac{p(z_l)}{p(\boldsymbol{a}^{z_l})} \prod_{m=1}^{M} p(a_m^{z_l} \mid x), \tag{3}$$

Finally, mapping function $f : X \longrightarrow Z$ recognizes object z_x, as the one having the highest posterior:

$$z_x = f(x) = \underset{z_l \in Z}{\operatorname{argmax}}\; p(z_l \mid x). \tag{4}$$

4. Evaluation of zero-shot learning for haptic recognition

So far, we presented the theoretical background of attribute-based ZSL. However, its use with haptic data collected by a robot is not trivial. The specific nature of haptic data and robotic limitations add challenges that require various adaptations of the aforementioned theoretical framework. First, haptic data collection is costly, requiring the optimization of training data. Second, both the choice of attributes and the nature of haptic data depend heavily on a robot's sensing capabilities, hindering comparisons and generic solutions. Third, data gathered from the robot can be multimodal (e.g. joints, temperature…) requiring processing and fusion techniques. Fourth, the spatial limitation of robot exploration leads to sparse and noisy, or missing data, which is challenging to the recognition algorithm. To assess the use of ZSL on haptics in the presence of all these difficulties, we started by applying it to the state of art PHAC-2 dataset [1].

The performance of a recognition algorithm may be evaluated by its classification accuracy, computation speed and memory requirement. Here we concentrate on classification accuracy as we are using an offline recognition

method and are not constrained by the training response time. Furthermore, we suppose that enough resources are available to run our algorithms. This was the case throughout our experiments which were carried out successfully on a PC with an Intel(R) Core(TM) i7-3840QM processor having a speed of 2.8 GHz and a RAM of 8 GB.

4.1. PHAC-2 database description

The PHAC-2 dataset[5] was designed to study how a robot can learn to describe its haptic perception using words. The developers of PHAC-2 use *haptic adjectives* which are binary high-level physical properties of objects such as 'hard', 'absorbent' and 'bumpy'. By considering the PHAC-2 adjectives as attributes, we can apply an attribute-based approach to perform ZSL.

Several characteristics motivated us to use PHAC-2 to evaluate our ZSL framework. First, it is probably the single largest object database with objects labeled by their haptic characteristics [5, 45, 46]. Second, the objects in the database encompass a variety of physical properties in terms of texture, material and stiffness. Third, the adjective definition and the binary associations between objects and adjectives have been developed using an arguably unbiased procedure. And finally, PHAC-2 provides data from multiple explorations of the same object, allowing for the testing of algorithms for robustness to inter-trial variabilities.

The PHAC-2 database contains 60 objects, labeled by a broad range of material, texture and stiffness related adjectives. Twenty four haptic adjectives were defined by human volunteers for the set of objects. For our analysis we utilized 19 adjectives (after removing adjectives present in less than 3 objects). The adjectives used by us were $A = \{absorbent, bumpy, compressible, cool, fuzzy, hard, hairy, metallic, porous, rough, scratchy, slippery, smooth, soft, solid, springy, squishy, textured, thick\}$.

In addition to the adjectives, the PHAC-2 database provides 48-dimensional vectors of raw haptic signals. These have been obtained using the Willow Garage PR2 robot[6] gripper, equipped with two BioTacs sensors[7]. The gripper provides the following kinesthetic information: aperture X_g and height Z_g in the robots torso coordinate frame. Additionally, each BioTac provides cutaneous information: core temperature T_{DC}, heat flow T_{AC}, static pressure P_{DC}, dynamic pressure P_{AC} and the voltage change for each of the 19 impedance sensing electrodes $[E_1 \dots E_{19}]$. These data items (X_g, Z_g and 23 signals from each BioTac) sum up to define the 48-dimensional vector. Ten vectors are provided for each object, one corresponding to each trial that had four *exploration procedure*, or EP (including a squeeze, hold, slow slide and fast slide on the object), making a total of 600 samples (60 objects $\times$ 10 trials).

4.2. Splitting the object set O

To perform ZSL using the PHAC-2 dataset, we split the 60 objects into two disjoint sets. To ensure that the results are independent from the chosen splits, we generated 5 splits $\{(Y_s, Z_s),\ s = 1, \dots, 5\}$ where each pair respects two constraints: $Y_s \cup Z_s = O$ and $Y_s \cap Z_s = \emptyset$. To generate each pair (Y_s, Z_s), we randomly chose for the test set 10 objects out of the 60: $Z_s = \{o_k \in O\ ,\ k \in rand(10, 60)\}$[8]. The remaining 50 objects were used for training: $Y_s = O - Z_s$. However, some objects have the same attributes vector and since ZSL identifies each test object by its attributes vector, we verified that the attributes vectors of all objects in Z_s were mutually different. If not, the random selection (of the 10 objects) was repeated until the condition was satisfied.

4.3. Feature Extraction

As in [1], we considered the minimum gripper aperture, mean aperture and travelled distance for each exploration as features. This results in a vector of 12 kinesthetic features (3 features $\times$ 4 EP). For the BioTac readings, we exploit the results from [26, 33, 47]. First, we removed the baseline activity from the BioTac readings by subtracting the mean of the first 100 readings. Next, the average of each signal was computed over the exploration time. The computed features were then normalized to have zero mean and standard deviation of 1. Since the BioTacs are not calibrated in the dataset, we considered each BioTac separately. Thus, by concatenating features of the 4 EP, we obtained a vector of 92 features for each BioTac (23 BioTac features $\times$ 4 EP). This vector was concatenated with the gripper's 12

[5] `https://github.com/IanTheEngineer/Penn-haptics-bolt`

[6] `http://www.willowgarage.com/pages/pr2/overview`

[7] `https://www.syntouchinc.com/sensor-technology/`

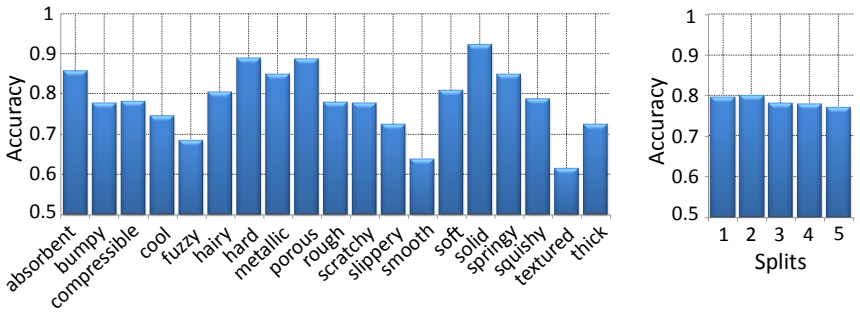
[8] $rand(k, n)$ returns k random numbers in range $1, \dots, n$.

features, which resulted in a vector of 104 features for each BioTac. This high dimension was reduced using Principal Component Analysis (PCA) to a 25-dimensional vector that justifies 95% of the variance. To sum up, an exploration trial results in 2 vectors of haptic features $x_{b1}, x_{b2} \in \mathbb{R}^{25}$, one for each BioTac.

4.4. Attributes classification

Attributes classification aims at estimating the presence of attribute a_m for a given feature vector x, i.e., to derive $p(a_m|x)$. By assuming attributes independence, we trained a Support Vector Machine (SVM) classifier for each attribute and each BioTac. When a test sample x is given to a SVM classifier, it returns a score $s(x)$ corresponding to the distance from x to the decision boundary. The attribute posterior $p(a_m|x)$ is estimated by transforming $s(x)$ into a probability using a sigmoid function.

We trained a binary classifier for attribute m and BioTac b_i using the training set $D^m_{train,b_i} = \{(x_{b_i}, a^y_m)$ s.t. $(x, y) \in D_{train}\}$. First, by analyzing the provided matrix $\boldsymbol{K}$, we noticed that for each attribute the number of objects in which it is present was significantly smaller than the number of objects from which it was absent. This leads to an unbalanced training set D^m_{train,b_i}: the ratio of class 1 samples to class 0 samples is $a : b$ with $a << b$. An unbalanced training set can lead to over-fitting of the over-represented class. To cope with this problem, we under-sampled class 0 by randomly removing pairs $(x_{b_i}, a^{y_k}_m)$ having $a^{y_k}_m = 0$ from D^m_{train,b_i} until we obtained an equal number of training samples for classes 0 and 1. This under-sampling was repeated multiple times, and sets giving the best results were used. Using the balanced set, we then trained a non-linear SVM classifier with a Gaussian kernel using the python machine learning library *Scikit-learn*[9]. Parameters C and γ were tuned using leave-one-out cross-validation, by varying C between 1 and 10^2 and γ between 10^{-2} and 1.



(a) (b)

Figure 2: Recognition accuracy of attributes binary classification.

Fig. 2a shows the average classification accuracy for each attribute. We averaged the accuracies obtained from the two BioTacs classifiers that were tested on the 5 different test object sets. The results show that attributes classifiers were able to learn from training data whether an attribute is present or absent in a novel object. However, performance varied from one attribute to the other: we obtained an accuracy of 92% for attribute *solid* whereas we obtained modest, though still better than random, performance for attributes *smooth* (64%) and *textured* (64%). To evaluate the influence of the object splits, we plot in Fig. 2b the attributes classification accuracies for each of the five splits (Sect. 4.2), averaged on the 2 BioTac and on all attributes. The figure shows that classification accuracies are not heavily influenced by the choice of object splits: the average accuracy is 78%, with a maximum difference of only 3% between splits 2 and 5.

While above we report the choice of attribute classification assuming attribute independence, we did also consider attribute classification without the assumption of independence. For this, we trained a multi-label SVM classifier for each BioTac b_i using $D_{train,b_i} = \{(x_{b_i}, \boldsymbol{a}^y)$ s.t. $(x, y) \in D_{train}\}$, with the Python *scikit-multilearn*[10] library implementing the classifier proposed in [48], and tuned the SVM parameters C and γ using a leave-one-out cross validation. This

[9] `http://scikit-learn.org/stable/`
[10] `http://scikit.ml/`

multi-label classifier predicts, given a test sample x, a vector of posteriors, one for each attribute. However, with this multi-label classifier, we obtained an average recognition accuracy of 0.31, which is lower than that 0.39 obtained assuming independent attributes (see table 1).

4.5. Recognition of novel objects

The final step consists of combining the attribute posteriors to infer the identity of the distinct test sample x (see Sect. 3.3). Since we distinguish features from each BioTac: $x = (x_{b1}, x_{b2})$, attributes posteriors $p(a_m^{z_l}|x_{b1})$ and $p(a_m^{z_l}|x_{b2})$ must be combined to infer $p(a_m^{z_l}|x)$. For this, we propose two methods. The first (that we name MAXDAP) considers only the highest posterior, by assuming that it is the most confident value between the two:

$$p(a_m^{z_l}|x) = \max(p(a_m^{z_l}|x_{b1}), p(a_m^{z_l}|x_{b2})). \tag{5}$$

The second method (AVGDAP) considers the average of both posteriors to eliminate the influence of misclassification:

$$p(a_m^{z_l}|x) = \frac{p(a_m^{z_l}|x_{b1}) + p(a_m^{z_l}|x_{b2})}{2}. \tag{6}$$

Attributes classification posteriors are used by DAP to perform zero-shot recognition according to equation (3). First, we compare in table 1 the use of MAXDAP and AVGDAP to combine BioTac posteriors by assuming uniform attributes posteriors. Since we use the MAP estimator in equation (4) to distinguish the ten test objects, the random classification accuracy for the ZSL algorithm is 0.1. The reported accuracies were therefore significantly higher than random chance. Performances of the two methods are similar, since the objects have homogeneous properties on their surfaces, giving close attributes posteriors for the two BioTacs.

Split	MAXDAP	AVGDAP
1	0.48	0.48
2	0.38	0.36
3	0.38	0.38
4	0.35	0.35
5	0.35	0.36
Average	0.39	0.39

Table 1: ZSL accuracy for independent attributes classifiers.

We compared three methods for measuring the attribute prior $p(a_m^{z_l})$:

1. based on its presence in training objects: $p(a_m = 1) = \frac{1}{N}\sum_{n=1}^{N} a_m^{y_n}$,
2. based on its presence in test objects: $p(a_m = 1) = \frac{1}{L}\sum_{l=1}^{L} a_m^{z_l}$,
3. based on a uniform distribution: $p(a_m = 1) = 0.5$.

We refer to these methods as *train-prior*, *test-prior* and *uni-prior* respectively. Note that test-prior is feasible because the attribute vectors of unknown objects are available to the robot in ZSL. We report in table 2, results obtained using AVGDAP method. Results show that the three methods give similar accuracies with a slight improvement with *uni-prior*. This is coherent with results reported in [14], which show that attribute prior estimation is not crucial.

4.6. Robustness to the choice of objects

Considering that recognized objects are experienced for the first time, the results listed in tables 1 and 2 show that our ZSL can be used for haptic recognition. However, as one would expect, we found that the recognition results vary depending on the training and test set (see the difference in performance across the 5 splits). This is an interesting finding because Fig. 2 shows that attributes classification accuracies themselves do not differ between the splits. This discrepancy highlights the influence of attribute space, that is that object closer in attributes space are more difficult to distinguish from one another, and motivated us to test the robustness of our method with more challenging data sets, where the similarities between the training and test sets are minimal.

Split	uni-prior	test-prior	train-prior
1	0.48	0.24	0.36
2	0.38	0.33	0.28
3	0.38	0.39	0.32
4	0.35	0.39	0.38
5	0.35	0.33	0.29
average	0.39	0.34	0.33

Table 2: DAP accuracy for independent attributes classifiers.

Previously, we reported results for five random training/test splits. To minimize the similarities between the training and test data, and make the recognition more challenging for our ZSL algorithm, we considered our objects in a 2-dimensional representation of haptic similarities (Fig. 3). To obtain this representation, we used PCA to reduce the dimension of each trial feature vector from $\mathbb{R}^{25}$ to $\mathbb{R}^2$ (by considering the first two principal components to represent the objects). Following this, we averaged the 10 vectors corresponding to the 10 trials of each object so as to obtain one vector that represents the object in the 2-dimensional space.

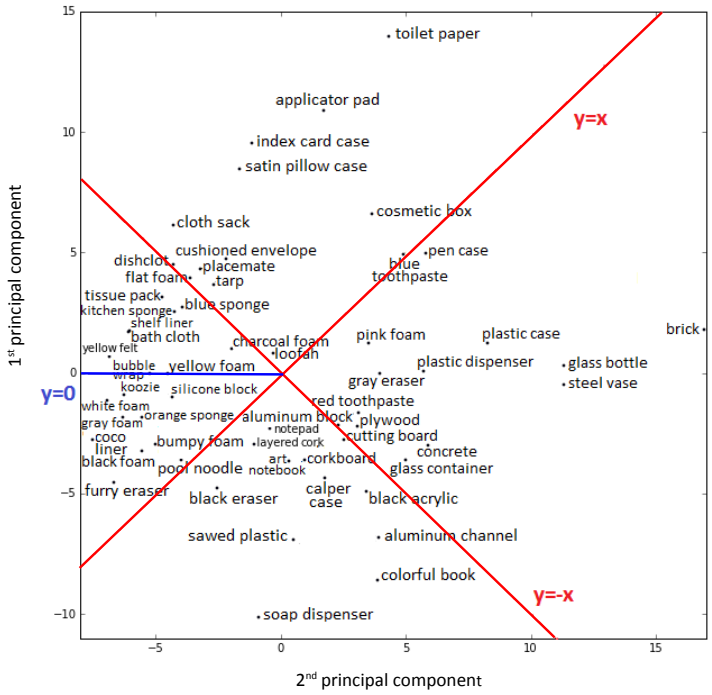


Figure 3: Splitting the PHAC-2 objects set into five partitions.

In this representation space, note that near objects feel similar when touched, and are harder to distinguish by a haptic recognition system. Therefore, to create challenging splits, we first used lines $y = x$ and $y = -x$ (red) and divided the object space into 4 partitions. Since the partition on the left of the graph is significantly denser than the others, we divided it again in two with line $y = 0$ (blue), to yield 5 partitions in total. We then considered 5 splits, by using each time one partition as test set and the 4 others for training. Since attribute-based ZSL cannot distinguish between objects having identical attribute vectors, in each test set we kept only objects having different attributes vectors. Next, to increase the attributes classifier's ability to generalize to new objects, we ensured that each attribute classifier is trained using at least 3 different objects. Overall, the above procedure helped us create challenging object splits that ensure that, a) test objects are *different* from training objects, and b), test objects are *similar*, making them

harder to distinguish.

In table 3, we report the classification accuracies with the challenging splits. Overall, we note a decrease in accuracies, as compared to the results obtained using random splits. This was expected since we intentionally made the ZSL task more difficult. Despite the decrease in accuracy, the recognition accuracy was above chance, highlighting the ability of the algorithm to recognize object very different from those it has trained on.

Split	number of objects	random accuracy	AVGDAP
1	11	0.09	0.18
2	7	0.14	0.21
3	7	0.14	0.3
4	10	0.1	0.22
5	11	0.09	0.27
Average			0.24

Table 3: DAP accuracy for challenging splits.

4.7. Influence of attribute numbers

Finally, we investigated the influence of the number of attributes on the ZSL accuracy. As mentioned before, we used 19 attributes to perform ZSL. To estimate the accuracy when using j attributes, we randomly picked up to 1000 combinations of j attributes out of the 19, and then averaged the accuracies of ZSL obtained with each of the combinations. We do so with j varying from 7 to 19. We started from 7 since it is the size of the minimum subset of attributes that allows to distinguish between objects. Results, reported in Fig. 4, show that, as expected, increasing the number of attributes generally improves the performance. Increasing the number of attributes improves the separation between objects, which are now represented in a higher dimensional space, and having more distinct attribute vectors between objects alleviates the influence of misclassifying them.

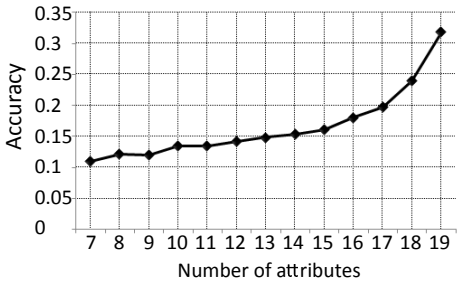


Figure 4: Recognition accuracy vs. the number of attributes utilized for ZSL.

4.8. Summary

To summarize, we developed a variant of ZSL for haptic recognition of novel objects. Using the PHAC-2 dataset, we analyzed the influence of several factors on the performance of the haptic ZSL, including the choice of attribute prior, the way of combining the attribute posterior, the number of attributes, and the object set split. Furthermore, we showed the robustness of our algorithm by minimizing the similarity between the training and test set. We show that even in the worst case, when the training and test objects are very different, the algorithm can still give above chance recognition accuracies.

Next, we applied the developed haptic ZSL on an anthropomorphic hand system, to test its performance in a real experimental setting.

5. Recognition of hand held daily life objects by a robotic hand

5.1. Motivation: Beyond PHAC 2

PHAC-2 offers an important amount of objects, attributes and haptic data allowing us to test the application of ZSL on haptics. However, the database was built in a controlled setting and with objects of regular shape, that may not always be the case in a real robotics application. Therefore, in our experimental setup we used a less controlled exploration, more realistic objects with heterogeneous surface properties, and different modalities of haptic data. The goal was to show how ZSL can be applied to real life robotic applications, and to recognize any new constraints the recognition system could face in such applications. Specifically:

1. We did not have a planned object exploration strategy (like in PHAC 2), and our object exploration is achieved during open-loop random grasps. Our robot grasped the object in an unknown position and orientation, and used whatever it senses as exploration data. This made the exploration short and coarse, but crucially more realistic.
2. We used daily-life objects with semantic meanings (e.g. bottle, mug and box). The objects were not of any particular shape. In addition, we allowed objects to have heterogeneous physical properties on their surfaces, meaning that the object can feel differently depending on the touched part. This requisites the exploration to be incremental and to include different sources of information.
3. We used a dexterous anthropomorphic robotic hand that offered not only cutaneous information, but also rich kinesthetic data, typical of whole hand object grasps.
4. We made use of available (online dictionary based) textual descriptions of the object in order to avoid the time-consuming human exploration process and minimize the human effort needed in this procedure. This is more suitable for real scenarios, since we aim at minimizing the cost of adding a new object to set O.

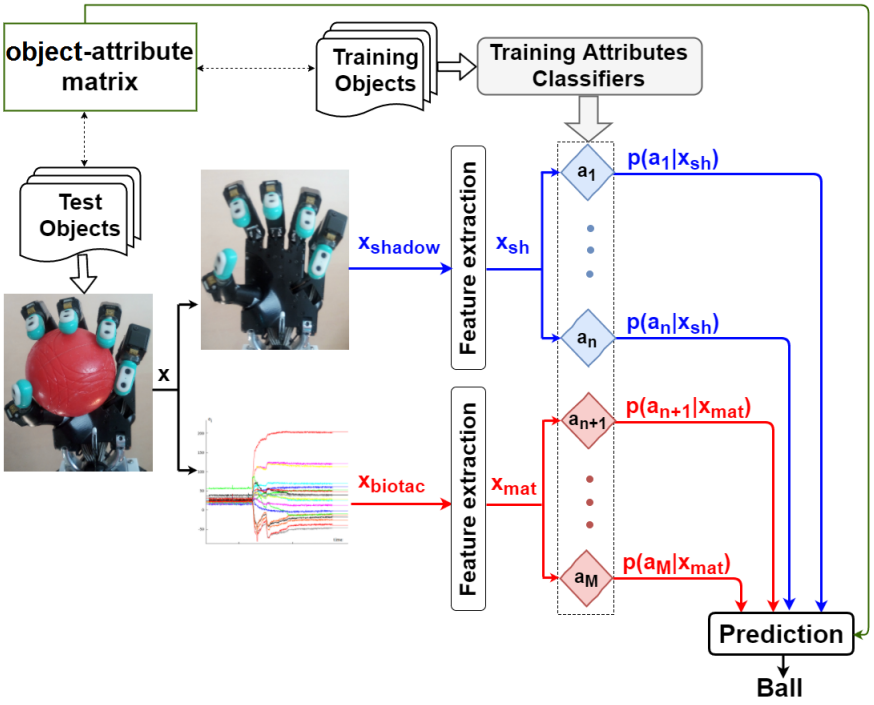


Figure 5: Our zero-shot recognition system overview.

In Fig. 5, we summarize the different steps of our robotic experimentation. First, we collected two disjoint sets of daily life objects, one for training and the other for testing. Then, we defined a list of attributes allowing us to describe both sets of objects, resulting in the object-attribute matrix $\boldsymbol{K}$. Next, we trained binary attribute classifiers:

using BioTac readings for material attributes and robot joint readings for shape attributes. Finally, haptic readings collected by exploring a new test object are introduced to the attributes classifiers and the resulting posteriors are used by the DAP model to infer the test object's identity.

5.2. Robot hand setup

Our robot setup consists of a cable-driven Shadow Dexterous Hand[11] with a BioTac sensor mounted on each fingertip. The encoders on the hand's 19 joints provide kinesthetic information $x_{shadow} = \{q_1, \ldots, q_{19}\}$, where q_i is the angular position of joint i, and each BioTac provides cutaneous information $x_{biotac} = \{T_{AC}, T_{DC}, P_{DC}, P_{AC}, E_1, \ldots, E_{19}\}$.

5.3. Daily life object set

We collected a set of different daily life objects to form our object set O. Our choice was based on multiple state of art studies that established lists of real objects that are interesting for robotic manipulation. The authors of [49] propose a list of objects ordered according to their relevance for automatic retrieval after surveying people with amyotrophic lateral sclerosis (ALS). A larger list of daily life objects categorized according to their use is given in [50]. The YCB database [51] also regroups a set of physical objects to be used for object manipulation benchmarking. We chose our objects from these previous works that were big enough to be grasped by the Shadow Hand, and which were not hot, or have sharp edges that could damage the BioTacs. Finally, we selected the 20 objects illustrated in Fig. 6 to form our object set.

Figure 6: Disjoint training (left) and test (right) object sets.

5.4. Attribute-based description

For attribute-based ZSL, the design of the dataset requires defining: (1) training objects Y, (2) test objects Z, (3) the set A of attributes that can be derived from the data collected with our setup and (4) matrix $\boldsymbol{K}$. First, we randomly split the object set O into two equal disjoint sets: $Y = \{$*cardboard box*, *glass bottle*, *plastic bottle*, *round container*, *mug*, *thermal mug*$\}$ and $Z = \{$*ball*, *rectangular container*, *tube*, *blender leg*, *bowl*, *glass*, *plastic cup*, *measuring cup*, *jar*, *salter*$\}$ as illustrated in Fig. 6. Next, we defined a set of attributes A appropriate for describing the haptic sensation of Y and Z. Here, we deal with a specific type of attribute which is semantic and binary-valued. Considering non-semantic [52] and real-valued attributes [40] is beyond the scope of this paper. Multiple works have studied the definition of attributes by relying on human expressing capabilities [1, 53] or linguistic knowledge databases [39, 38]. In PHAC-2, authors relied on human experiments to define the adjectives list, which can depend on the participants' choice of words. Here, we chose a more objective approach by making use of objects names to extract their textual descriptions from online dictionaries[12]. From these descriptions, we chose attributes that could be sensed by our robot. For instance, from the definition of a bottle: *a glass or plastic container that has a narrow neck and usually*

[11] shadowrobot.com

[12] We used merriam-webster.com and en.oxforddictionaries.com.

has no handle, we extracted the underlined statements as attributes. Overall, by analyzing all objects descriptions, we extracted a list of 11 shared attributes : $A = \{$*porcelain*, *plastic*, *glass*, *cardboard*, *steel*, *cylindrical*, *round*, *rectangular*, *concave*, *has a handle*, *has a narrow part*$\}$.

Next, we set $a_m^o = 1$ if the attribute a_m is used to describe the object o as a required property (e.g. *has a narrow part* for a bottle), and $a_m^o = 0$ if the attribute is an undesired or unnecessary property (e.g. *has no handle* for a bottle). Using this procedure we obtained the object-attribute matrix $\boldsymbol{K}$ illustrated in table 4. Training objects y_n having identical $\boldsymbol{a}^{y_n}$ are indicated only once in the table, this is why we have 6 instead of 10 training objects. Note that each test object z_l has a specific $\boldsymbol{a}^{z_l}$ that will be used to distinguish it from other objects during recognition.

	porcelain	plastic	cardboard	glass	steel	cylindrical	round	rectangular	concave	has handle	has narrow part
cardboard box	0	0	1	0	0	0	0	1	0	0	0
glass bottle	0	0	0	1	0	1	0	0	0	0	1
plastic bottle	0	1	0	0	0	1	0	1	0	0	1
round container	0	1	0	0	0	0	1	0	0	0	0
mug	1	0	0	0	0	1	0	0	1	1	0
thermal mug	0	1	0	0	1	1	0	0	1	1	0
ball	0	1	0	0	0	0	1	0	0	0	0
rectangular container	0	1	0	0	0	0	0	1	0	0	0
tube	0	1	0	0	0	1	0	0	0	0	0
blender	0	1	0	0	0	1	0	0	0	0	1
bowl	1	0	0	0	0	0	1	0	1	0	0
glass	0	0	0	1	0	1	0	0	1	0	0
plastic cup	0	1	0	0	0	1	0	0	1	0	0
measuring cup	0	1	0	0	0	1	0	0	1	1	0
jar	0	1	0	1	0	0	0	1	0	0	0
salter	0	0	0	1	1	1	0	0	0	0	1

Table 4: Class-attribute matrix $\boldsymbol{K}$ for training (upper) and test (lower) objects.

5.5. Data collection

The next step was data collection from the training objects, to learn attributes. We note that our attributes can be decomposed into 2 categories: material and shape attributes. We referred to the state of art on haptic exploration to choose the best procedure a hand must perform in order to recognize shapes and materials. According to [54], materials can be measured by performing a static contact between sensors and object surface. Shape can be inferred from the hand grasping/enclosing the object. Thus, we chose to explore each object by grasping/enclosing, which combines the exploration procedures required to perceive both, material and shape.

During the training phase, each training object was handed to the Shadow hand by an experimenter 10 times in random positions and orientations. To grasp the object, the robot fingers were spread out to their joint limits. Then, they were closed by setting a desired constant current to each joint actuator. The currents (and corresponding cable tensions) were kept low enough to avoid damage during contact with object, or in the case of self collisions, while ensuring a "good" contact between the BioTac and object. Once all joints stopped (either because the finger contacted the object, or because it reached its joint limit), the contact was maintained for 20 seconds to obtain the thermal equilibrium between sensors and object. Different examples are illustrated in Fig. 7.

Since our attributes have different nature, we assume attribute independence and learn each one separately. From each grasp, we gathered BioTac readings x_{biotac} serving as inputs for material attributes classifiers, and encoders

Figure 7: Examples of object grasps by the Shadow Hand.

readings x_{shadow} serving as inputs for shape attributes classifiers. After exploring all objects from Y, the collected data were used to build a training set $D^{a_m}_{train}$ for each shape attribute and $D^{a_m}_{train,b}$ for each material attribute-BioTac pair.

Each test object was grasped up to 15 times, again using the same grasping procedure as during training, to build the test set D_{test}. Grasps in which none of the fingers touched the object were dropped as they do not include material information. A video of the experiments is provided with this paper[13].

5.6. Attributes classification

For material attributes, we chose the same feature extraction technique as for the previous experimental setup (see Sect. 4.3). We used a time average of the features and linear SVM as suggested in [26]. But in contrast to that work, since we performed a static contact, we did not consider the vibrations signal P_{AC} from the BioTac. We obtained a feature vector of 22 normalized means which was reduced using PCA to a 8-dimensional vector x_{mat} that explained more than 98% of the variance. The resulting x_{mat} was used for the classification of the material attributes.

For each shape attribute, a binary classifier was trained using joints position measurements from the Shadow hand x_{shadow}. Joints that did not contribute to the closing procedure (e.g. the wrist) were excluded, resulting in a feature vector x_{sh} of 10 angular positions input to the shape attributes classifiers. The classification was performed using nonlinear SVM with a Gaussian kernel. We used the same sigmoid function as in Sect. 4.4 to convert the SVM classification score to an attribute posterior.

5.7. DAP – handling robotic constraints

To classify a test sample x, we introduced it in each attributes binary classifier, to obtain a set of posteriors $\{p(a_m \mid x), m = 1, \ldots, 19\}$. The attributes posteriors and object-attribute matrix are used by DAP (see Sect. 3.3) to return the object class.

When the hand grasps an object, it provides a data sample $x = (x_{sh}, x_{mat})$ where $x_{mat} = (x_{1,mat}, \ldots, x_{B,mat})$ with $1 \leq B \leq 5$ depending on the number of BioTacs in contact. Thus, the attributes classifiers return $p(a_m \mid x_{sh})$ for shape attributes and $\{p(a_m \mid x_{b,mat}), b = 1, \ldots, B\}$ for material attributes. These posteriors must be combined to infer the final attribute posterior required in equation (3).

Furthermore, the attribute posteriors have to take into account several constraints posed by our realistic experimental setup, which were absent when we worked with the PHAC-2 database:

1. Test objects can be made of multiple materials. Thus, BioTacs on different fingers can be in touch with different materials (e.g. the salter grasp in Fig. 7), requiring the integration of information from different fingers to estimate the material attributes.
2. Objects can be heterogeneous and a grasp may provide only a local view. For example, in Fig. 7 (right) the hand grasps the lower part of the bottle and misses the presence of the narrow neck. Therefore, we need to deal with information missing from touched parts and to combine grasps to obtain a global view.
3. The number of touching fingers B can vary from one grasp to another, giving a different size of the test sample each time, which must be taken into account in multi-grasp classification.

In the next sections, we propose and test different solutions proposed to take into account these constraints.

[13]The video can also be found on the IDH YouTube channel: https://youtu.be/Ekd28b0BiQs

5.7.1. Single grasp classification – Combining information from various BioTacs

Since the test object is grasped in an unknown pose, the number of BioTacs B making contact may vary for each grasp. The contact of a BioTac is detected when the difference in static pressure exceeds a given threshold. For each material attribute a_m, we obtain a set of posteriors from the contacting BioTac classifiers $\{p(a_m \mid x_{1,mat}), \ldots, p(a_m \mid x_{B,mat})\}$. To obtain the final attribute posterior for the material attributes, we tested both MAXDAP and AVGDAP (see Sect. 4.5) to combine classifications from the contacting BioTacs. MAXDAP considers only the BioTac that is most confident about the presence of the a_m, and was implemented as:

$$p(a_m = 1 \mid x) = \max_{b=1,\ldots,B} p(a_m = 1 \mid x_{b,mat}). \tag{7}$$

AVGDAP on the other hand, combines all contacting BioTacs by averaging their posteriors:

$$p(a_m = 1 \mid x) = \frac{1}{B} \sum_{b=1}^{B} p(a_m = 1 \mid x_{b,mat}). \tag{8}$$

Finally, while the material attribute posteriors were obtained using either MAXDAP or AVGDAP, the shape attributes were assessed as: $p(a_m = 1 \mid x) = p(a_m = 1 \mid x_{sh})$.

5.7.2. Single grasp classification – Handling local view

In regard to both material and shape attributes, a grasp may miss some attributes that are not present in the touched part of the object. This implies that if $p\,(a_m = 1 \mid x) < 0.5$, the attribute could be absent from the whole object or only from that particular grasp. To alleviate the effect of a possible misclassification, we replace the attribute posterior with a uniform distribution if the attribute is absent from x. For shape attributes:

$$p(a_m = 1 \mid x) = max(0.5, p(a_m = 1 \mid x_{sh})), \tag{9}$$

and for material attributes:

$$p(a_m = 1 \mid x) = \frac{1}{B} \sum_{b=1}^{B} max(0.5, p(a_m = 1 \mid x_{b,mat})). \tag{10}$$

We refer to this method as 'Local DAP' or LDAP.

5.7.3. Multi-grasp classification – Developing a global view

While single grasp classification can recognize objects by making some assumptions about absent attributes, combining several grasps is obviously advantageous as it gives a wider view of the object, and thus is expected to improve recognition performance [6]. Grasping an object T times in different positions results in a set $x = \{x^{(1)}, \ldots, x^{(T)}\}$ of test samples. We compare two approaches to exploit information from multiple grasps: data fusion and decision fusion.

Data fusion merges data from the T grasps to form one "super grasp", that can be used to classify an object just like a single grasp. Formally, for shape attributes, we have:

$$p(a_m = 1 \mid x) = \frac{1}{T} \sum_{t=1}^{T} max(0.5, p(a_m = 1 \mid x_{sh}^{(t)})), \tag{11}$$

and for material attributes:

$$p(a_m = 1 \mid x) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{B_t} \sum_{b=1}^{B_t} max(0.5, p(a_m = 1 \mid x_{b,mat}^{(t)})). \tag{12}$$

We refer to this method as 'Data Fusion for Multi-grasp DAP ' or DF-MDAP.

Decision fusion separately classifies each grasp in the sequence starting from $x^{(1)}$ until k similar classifications are made or until $x^{(T)}$ is reached. The classification that reaches k is taken as final decision on the object. We refer to this method as 'Similar classifications for Multi-grasp DAP' or SC-MDAP.

5.8. Experimental evaluation and results

We made a series of experiments to evaluate the multiple adaptations of DAP to our experimental setup. For better analysis, we decomposed the test set Z into 3 subsets: objects having homogeneous material and shape properties Z_{hom} ={*ball*, *rectangular container*, *tube*}, objects whose shape properties can differ according to the touched part $Z_{het,sh}$ ={*blender leg*, *bowl*, *glass cup*, *plastic cup*, *measuring cup*} and objects made of multiple materials: $Z_{het,mat}$ ={*jar*, *salter*}.

5.8.1. Attributes classification

First, we evaluate the performance of the binary classification of attributes. In Fig. 8, we present the accuracies of attributes classifiers on the test set. All accuracies are averaged across the test trials. Material attributes are additionally averaged over the five BioTacs. Overall, attribute classification achieved a satisfying average accuracy of 89%. However, attribute *plastic* had an accuracy of 45% which is considerably lower than the rest. This was probably due to the variety of plastic types used in the training and test sets. For example, the round container is made of a softer plastic than the blender leg.

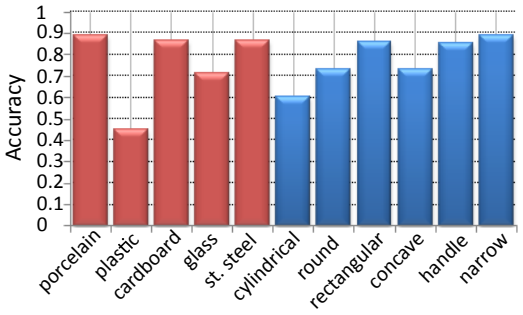


Figure 8: Recognition accuracy of attributes binary classification.

To analyze the performance of attributes classifiers on each object, we averaged $p(a_m = 1 \mid x)$ for each object-attribute pair across the test trials of each object. Ideally, if x is collected from object z_l then $p(a_m = 1 \mid x)$ should be close to $a_m^{z_l}$ since $p(a_m = 1 \mid x) \sim 1$ if $a_m^{z_l} = 1$ and $p(a_m = 1 \mid x) \sim 0$ otherwise. Matrix $\boldsymbol{S}$ presented in Fig. 9 measures L_1 distance between attributes binary labels and their posteriors:

$$\boldsymbol{S}_{l,m} = \| a_m^{z_l} - \frac{1}{\mid D_{test}^{z_l} \mid} \sum_{x \in D_{test}^{z_l}} p(a_m = 1 \mid x) \|_1, \tag{13}$$

$D_{test}^{z_l}$ represents the set of test samples gathered from object z_l. We note that the majority of distances are lower than 0.5, indicating that objects are well classified by the corresponding attribute classifier. However, some classifiers perform badly on some objects, probably because these are *too* different from the ones they have trained on. For instance, the *plastic* attribute column presents high distances for the ball, rectangular container and jar because of the difference between the plastic material constituting these objects, and that constituting the training objects.

5.8.2. Single grasp DAP

Recognition accuracies of test objects using a single grasp are presented in table 5. First, it is clear that the accuracy obtained with homogeneous objects is better than that obtained with heterogeneous ones. This is because all object properties can be felt from a single grasp. Furthermore, all BioTacs touch the same material over a homogeneous object, "collaborating" to give a more confident classification. However, this collaboration becomes more delicate for objects belonging to $Z_{het,mat}$, which explains the accuracy deterioration. Moreover, we note that AVGDAP performs better than MAXDAP for heterogeneous objects. This is probably because AVGDAP averages the decision from the BioTacs, and is hence less sensitive to errors from individual sensors.

However, LDAP outperforms all methods because we found that more often than not, random grasps on daily life objects lead missing some parts of the object, and the LDAP can deal efficiently with absent attributes.

	porcelain	plastic	cardboard	glass	st. Steel	cylind.	round	rectang.	concave	handle	narrow
ball	0.38	0.61	0.09	0.25	0.16	0.33	0.58	0.20	0.02	0.29	0.22
rec. cont.	0.21	0.73	0.33	0.14	0.21	0.43	0.15	0.35	0.05	0.11	0.11
tube	0.11	0.36	0.14	0.48	0.19	0.08	0.29	0.10	0.06	0.02	0.25
blender	0.19	0.54	0.12	0.27	0.25	0.60	0.01	0.33	0.37	0.06	0.33
bowl	0.17	0.38	0.01	0.31	0.12	0.13	0.60	0.19	0.51	0.25	0.17
glass	0.20	0.35	0.07	0.38	0.09	0.50	0.04	0.11	0.59	0.06	0.32
plas. cup	0.05	0.34	0.06	0.46	0.20	0.50	0.02	0.09	0.50	0.01	0.56
msr. cup	0.14	0.47	0.21	0.37	0.05	0.46	0.08	0.15	0.67	0.76	0.23
jar	0.12	0.70	0.24	0.67	0.21	0.59	0.15	0.45	0.06	0.03	0.16
salter	0.14	0.29	0.08	0.62	0.62	0.38	0.06	0.31	0.11	0.01	0.21

Figure 9: Distance between attributes binary labels and their posteriors for test objects.

Method	Z_{hom}	$Z_{het,sh}$	$Z_{het,mat}$	Z
MAXDAP	**77.78%**	44.83%	20%	52.29%
AVGDAP	**77.78%**	50%	33.33%	56.88%
LDAP	75%	**55.17%**	**40%**	**59.63%**

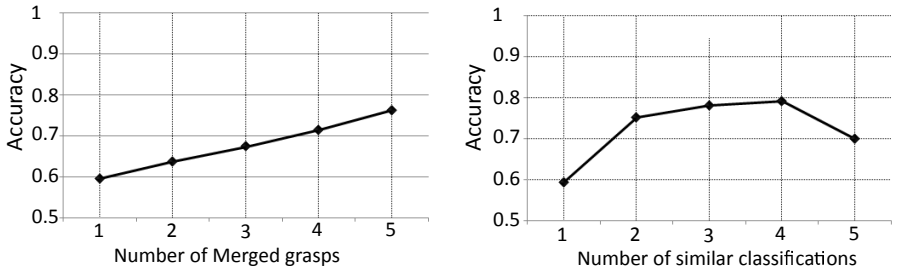
Table 5: Recognition accuracy of DAP on test objects from a single grasp.

Note that table 5 should however be viewed considering the fact that single grasps, by nature, are constrained by local object properties. Consider the three example grasps on the measuring cup during the single grasp classification (Fig. 10). We found that for 83% of grasps having the fingers touching the cylindrical part, the measuring cup is classified as a *tube*. All the grasps touching the upper side lead to the measuring cup being classified as the *pink cup*, and 66% of grasps touching the handle are classified as a measuring cup. These cases should not all be considered as misclassifications because this is the best decision that can be made from the given local grasp. For instance, a human touching the cylindrical part of the measuring cup can also not be sure if this is the tube, the pink cup or the measuring cup, since the three objects share the same local shape. Hence, table 5 shows the limitation of single grasps for object recognition.

Figure 10: Representative grasps on the "measuring cup" show how different grasps may indicate distinct kinematic features.

5.8.3. Multi-grasp DAP

Using a single grasp, LDAP gave the best recognition accuracy of 59.63% over all objects, which is modest, hence encourages multi-grasp recognition. Next, we checked the recognition accuracies during multiple grasps, using both DF-MDAP and SC-MDAP. In Fig. 11, we analyzed the DF-MDAP and SC-MDAP performances by changing their parameters T and k, respectively. We note that DF-MDAP accuracy increases constantly with the number of merged grasps T. This is understandable as an increase of T improves the information contained in the “super grasp". Since LDAP gave the best accuracy, we used it to perform each single grasp classification for SC-MDAP. For SC-MDAP, the best accuracies were obtained between $k = 2$ and $k = 4$. The presence of an *optimal* k can be explained by the fact that,



(a) (b)

Figure 11: Recognition accuracy vs. number of combined grasps T for DF-MDAP (a) and vs. number of similar classifications k for SC-MDAP (b).

while increasing the number of required similar classifications helps removing noisy classifications, it becomes harder to find k similar classifications when k becomes large. Overall, both methods improve the performance compared to single grasp recognition ($k = 1, T = 1$).

		ball	rect. cont.	tube	blender	bowl	glass	plas. cup	msr. cup	jar	salter
Z_{hom}	ball	0.6	0.1	0.2	0	0.1	0	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
$Z_{het,sh}$	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
$Z_{het,mat}$	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

(a) LDAP

ball	rect. cont.	tube	blender	bowl	glass	plas. cup	msr. cup	jar	salter
0.85	0	0.15	0	0	0	0	0	0	0
0	0.86	0.14	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0.99	0	0	0.01	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0.67	0.28	0	0.04	0	0	0	0
0	0	0.9	0	0	0.03	0.04	0.03	0	0
0	0.22	0.61	0	0	0	0	0	0.17	0
0	0	0	0	0	0	0	0	0	1

(b) DF-MDAP

ball	rect. cont	tube	blender	bowl	glass	plas. cup	msr. cup	jar	salter
1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0
0	0	0.8	0	0	0	0.2	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1

(c) SC-MDAP

Figure 12: Confusion matrices of LDAP , DF-MDAP ($T = 5$) and SC-MDAP ($k = 4$) respectively. SC-MDAP achieved perfect recognition of 8 out of 10 new objects.

Figure 12 shows the confusion matrices with different DAP choices. The best classifications were achieved with SC-MDAP. We note that multi-grasp recognition was very beneficial to correct Z_{hom} misclassifications, to the point

that SC-MDAP reached an accuracy of 100% for all objects. For $Z_{het,sh}$, we note that the blender, bowl and glass are correctly classified by both multi-grasp methods. The plastic cup is still frequently confused with a tube by DF-MDAP but these misclassifications were overcome with SC-MDAP. However, all methods performed weakly on the measuring cup and on the jar, which were often confused with the tube. This was because the attributes classifications were poor on these objects (see distance matrix in Fig. 9), subsequently influencing their classification.

Overall we could analyze and develop a haptic ZSL algorithm for an anthropomorphic hand that enables very good recognition of daily life objects that the robot encounters for the first time (see Fig. 12c).

6. Discussion and Conclusion

In this work, we proposed a haptic zero shot learning algorithm that enables a robot to recognize novel objects, which it has never been trained on before. First, we applied ZSL to probably the best haptic object database for ZSL (the PHAC-2), which includes 60 objects with a wide variety of texture, material and stiffness properties. This allowed us to analyze the possibilities and constraints associated with the application of ZSL to haptic recognition. We used this analysis to understand if and how the choice of attribute classifier (Sect. 4.4), the way of combining attribute posteriors (Sect. 4.5), the choice of training set (Sect. 4.6) and the number of attributes (Sect. 4.7) affect recognition performance. The algorithm developed from this analysis was then applied on an anthropomorphic robot hand, to make it recognize unexplored objects (Sect. 5). In this realistic scenario, we further optimized our algorithm and enabled it to account for heterogeneous objects (Sect. 5.7) and to integrate information from multiple grasps (Sect. 5.7.3). Our final algorithm enabled the robot to correctly recognize eight out of ten objects, that it grasps for the first time (third panel of Fig.12).

Note that the recognition rate in our real robot experiment (Fig.12) was much better than in the PHAC-2 database recognition (Tables 1 and 2). Comparison is difficult, as different objects were used in the two cases. However, it is still interesting to note that the real robot implementation worked better, considering that it used coarse grasps to explore heterogeneous objects of different shapes, compared to recognition with the PHAC-2 database in which objects were homogeneous, regularly shaped, and explored using a regular and well controlled procedure. However, we believe that the results of our robot experiment are in fact a better indicator of the capabilities of our algorithm for haptic object recognition. This is because of several reasons. Primarily, the PHAC-2 database focuses on the tactile properties of objects, and not on shape (all objects were in fact of the same shape in the dataset). On the other hand, shape is definitely a fundamental feature for haptic recognition during grasps, and our robot is able to integrate both tactile and shape information (through its joint angles) efficiently to improve object recognition. Moreover, while homogeneous objects (as in PHAC-2) may intuitively feel to be easier to recognize, this may not always be true. Most real life objects are heterogeneous and in fact the heterogeneous nature of an object, if explored well (we propose one way in Sect. 5.5), can act as a signature for the object, making it easier to recognize. This was probably the case in our robot experiment as well. Finally, the recognition of the two discussed features, i.e material heterogeneity and shape, are further improved by our algorithm by using multiple grasps in our robot experiment. In PHAC-2 on the other hand, the homogeneity of objects and regularity of explorations make the data similar across exploration trials, making simulated multiple grasps irrelevant.

The results obtained with our robotic setup show the promising capabilities of haptic ZSL for object recognition. This is very encouraging, given that this is still a prototype system that can be improved in several aspects. First, in this work we focused on the ZSL problem by assuming that all evaluated objects have never been experienced before. However, robots encounter previously explored as well as unexplored objects, and the integration of ZSL with other suggested multi-class object recognition systems can enable them to recognize known objects and progressively integrate novel ones. Second, since the goal of this study was to introduce and adapt ZSL to haptics, we adopted a simplistic approach using semantic binary attributes and the DAP model. Semantic representation and learning can be improved by investigating new techniques as those presented in [55]. Finally, in this work we study haptic object recognition by assuming that vision is unavailable, with the prospect in the future of combining the results with visual object recognition to improve general object recognition by machines [56]. In summary, our work highlights the efficiency of ZSL for haptic recognition, and we hope it will provide a start to works for optimizing haptic object recognition by autonomous robots in uncertain environments.

Acknowledgement

We would like to thank the authors of [1] for providing us with the data and the support for using the PHAC-2 database. Zineb Abderrahmane is supported by the Ministry of Higher Education and Scientific Research of Algeria through the Excellence Fellowship.

References

[1] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, T. Darrell, K. J. Kuchenbecker, Robotic learning of haptic adjectives through physical interaction, Robotics and Autonomous Systems 63 (2015) 279–292.
[2] K. Grauman, B. Leibe, Visual object recognition, Synthesis lectures on artificial intelligence and machine learning 5 (2) (2011) 1–181.
[3] M. Johnsson, C. Balkenius, Haptic perception with self-organizing anns and an anthropomorphic robot hand, Journal of Robotics 2010.
[4] M. Madry, L. Bo, D. Kragic, D. Fox, St-hmp: Unsupervised spatio-temporal feature learning for tactile data, in: Robotics and Automation (ICRA), 2014 IEEE International Conference on, IEEE, 2014, pp. 2262–2269.
[5] T. Corradi, P. Hall, P. Iravani, Bayesian tactile object recognition: learning and recognising objects using a new inexpensive tactile sensor, in: Robotics and Automation (ICRA), 2015 IEEE International Conference on, IEEE, 2015, pp. 3909–3914.
[6] A. Schmitz, Y. Bansho, K. Noda, H. Iwata, T. Ogata, S. Sugano, Tactile object recognition using deep learning and dropout, in: IEEE-RAS Int. Conf. on Humanoid Robots, 2014, pp. 1044–1050.
[7] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, W. Burgard, Object identification with tactile sensors using bag-of-features, in: Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on, IEEE, 2009, pp. 243–248.
[8] U. Martinez-Hernandez, N. F. Lepora, T. J. Prescott, Active haptic shape recognition by intrinsic motivation with a robot hand, in: World Haptics Conference (WHC), 2015 IEEE, IEEE, 2015, pp. 299–304.
[9] H. Liu, D. Guo, F. Sun, Object recognition using tactile measurements: Kernel sparse coding methods, IEEE Trans. on Instrumentation and Measurement 65 (3) (2016) 656–665.
[10] X. Yu, Y. Aloimonos, Attribute-based transfer learning for object categorization with zero/one training example, Computer Vision–ECCV 2010 (2010) 127–140.
[11] M. Elhoseiny, B. Saleh, A. Elgammal, Write a classifier: Zero-shot learning using purely textual descriptions, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2584–2591.
[12] C. H. Lampert, H. Nickisch, S. Harmeling, Attribute-based classification for zero-shot visual object categorization, IEEE Trans. on Pattern Analysis and Machine Intelligence 36 (3) (2014) 453–465.
[13] Y. Cheng, X. Qiao, X. Wang, Q. Yu, Random forest classifier for zero-shot learning based on relative attribute, IEEE Transactions on Neural Networks and Learning Systems.
[14] C. H. Lampert, H. Nickisch, 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.
[15] P. K. Allen, K. S. Roberts, Haptic object recognition using a multi-fingered dextrous hand, in: Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on, IEEE, 1989, pp. 342–347.
[16] S. Caselli, C. Magnanini, F. Zanichelli, Haptic object recognition with a dextrous hand based on volumetric shape representations, in: Multisensor Fusion and Integration for Intelligent Systems, 1994. IEEE International Conference on MFI'94., IEEE, 1994, pp. 280–287.
[17] A. Bierbaum, K. Welke, D. Burger, T. Asfour, R. Dillmann, Haptic exploration for 3d shape reconstruction using five-finger hands, in: Humanoid Robots, 2007 7th IEEE-RAS International Conference on, IEEE, 2007, pp. 616–621.
[18] N. Gorges, S. E. Navarro, H. Wörn, Haptic object recognition using statistical point cloud features, in: Advanced Robotics (ICAR), 2011 15th International Conference on, IEEE, 2011, pp. 15–20.
[19] M. Meier, M. Schopfer, R. Haschke, H. Ritter, A probabilistic approach to tactile shape reconstruction, IEEE Transactions on Robotics 27 (3) (2011) 630–635.
[20] M. Jin, H. Gu, S. Fan, Y. Zhang, H. Liu, Object shape recognition approach for sparse point clouds from tactile exploration, in: Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on, IEEE, 2013, pp. 558–562.
[21] N. Gorges, S. E. Navarro, D. Göger, H. Wörn, Haptic object recognition using passive joints and haptic key features, in: Robotics and Automation (ICRA), 2010 IEEE International Conference on, IEEE, 2010, pp. 2349–2355.
[22] S. E. Navarro, N. Gorges, H. Wörn, J. Schill, T. Asfour, R. Dillmann, Haptic object recognition for multi-fingered robot hands, in: Haptics Symposium (HAPTICS), 2012 IEEE, IEEE, 2012, pp. 497–502.
[23] 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.
[24] A. J. Spiers, M. V. Liarokapis, B. Calli, A. M. Dollar, Single-grasp object classification and feature extraction with simple robot hands and tactile sensors, IEEE Trans. on Haptics 9 (2) (2016) 207–220.
[25] M. Johnsson, C. Balkenius, Recognizing texture and hardness by touch, in: Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, IEEE, 2008, pp. 482–487.
[26] J. Hoelscher, J. Peters, T. Hermans, Evaluation of tactile feature extraction for interactive object recognition, in: IEEE-RAS Int. Conf. on Humanoid Robots, 2015, pp. 310–317.
[27] Z. Su, J. A. Fishel, T. Yamamoto, G. E. Loeb, Use of tactile feedback to control exploratory movements to characterize object compliance, Frontiers in neurorobotics 6.
[28] J. Yang, H. Liu, F. Sun, M. Gao, Tactile sequence classification using joint kernel sparse coding, in: Neural Networks (IJCNN), 2015 International Joint Conference on, IEEE, 2015, pp. 1–6.
[29] A. Aggarwal, P. Kampmann, J. Lemburg, F. Kirchner, Haptic object recognition in underwater and deep-sea environments, Journal of field robotics 32 (1) (2015) 167–185.

[30] M. Klingensmith, M. Herrmann, S. S. Srinivasa, Object modeling and recognition from sparse, noisy data via voxel depth carving, in: Experimental Robotics, Springer, 2016, pp. 697–713.
[31] M. V. Liarokapis, B. Calli, A. J. Spiers, 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.
[32] H. Soh, Y. Demiris, Incrementally learning objects by touch: Online discriminative and generative models for tactile-based recognition, IEEE transactions on haptics 7 (4) (2014) 512–525.
[33] D. Xu, G. E. Loeb, J. A. Fishel, Tactile identification of objects using bayesian exploration, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2013, pp. 3056–3061.
[34] D. Tanaka, T. Matsubara, K. Ichien, 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.
[35] M. Kaboli, G. Cheng, Novel tactile descriptors and a tactile transfer learning technique for active in-hand object recognition via texture properties, in: IEE-RAS International Conference on Humanoid Robots-Workshop Tactile sensing for manipulation: new progress and challenges, 2016.
[36] M. Kaboli, R. Walker, G. Cheng, Re-using prior tactile experience by robotic hands to discriminate in-hand objects via texture properties, in: IEEE Int. Conf. on Robotics and Automation (ICRA), 2016, pp. 2242–2247.
[37] Y. Xian, C. H. Lampert, B. Schiele, Z. Akata, Zero-shot learning-a comprehensive evaluation of the good, the bad and the ugly, arXiv preprint arXiv:1707.00600.
[38] M. Rohrbach, M. Stark, G. Szarvas, I. Gurevych, B. Schiele, What helps where–and why? semantic relatedness for knowledge transfer, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp. 910–917.
[39] O. Russakovsky, F.-F. Li, Attribute learning in large-scale datasets., in: ECCV Workshops (1), Vol. 6553, 2010, pp. 1–14.
[40] D. Parikh, K. Grauman, Relative attributes, in: Computer Vision (ICCV), 2011 IEEE Int. Conf. on, 2011, pp. 503–510.
[41] P. Kankuekul, A. Kawewong, S. Tangruamsub, O. Hasegawa, Online incremental attribute-based zero-shot learning, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 3657–3664.
[42] F. X. Yu, L. Cao, R. S. Feris, J. R. Smith, S.-F. Chang, Designing category-level attributes for discriminative visual recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 771–778.
[43] D. Jayaraman, K. Grauman, Zero-shot recognition with unreliable attributes, in: Advances in Neural Information Processing Systems, 2014, pp. 3464–3472.
[44] B. Romera-Paredes, P. Torr, An embarrassingly simple approach to zero-shot learning, in: International Conference on Machine Learning, 2015, pp. 2152–2161.
[45] H. Gu, S. Fan, H. Zong, M. Jin, H. Liu, Haptic perception of unknown object by robot hand: Exploration strategy and recognition approach, International Journal of Humanoid Robotics 13 (03) (2016) 1650008.
[46] M. Regoli, N. Jamali, G. Metta, L. Natale, Controlled tactile exploration and haptic object recognition, in: Advanced Robotics (ICAR), 2017 18th International Conference on, IEEE, 2017, pp. 47–54.
[47] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, N. Fitter, J. C. Nappo, T. Darrell, et al., Using robotic exploratory procedures to learn the meaning of haptic adjectives, in: IEEE Int. Conf. on Robotics and Automation, IEEE, 2013, pp. 3048–3055.
[48] J. Read, B. Pfahringer, G. Holmes, E. Frank, Classifier chains for multi-label classification, Machine learning 85 (3) (2011) 333–359.
[49] Y. S. Choi, T. Deyle, T. Chen, J. D. Glass, C. C. Kemp, A list of household objects for robotic retrieval prioritized by people with als, in: 2009 IEEE International Conference on Rehabilitation Robotics, IEEE, 2009, pp. 510–517.
[50] K. Matheus, A. M. Dollar, Benchmarking grasping and manipulation: properties of the objects of daily living, in: IEEE/RSJ Int. Conf. on Robots and Intelligent Systems (IROS), IEEE, 2010, pp. 5020–5027.
[51] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, A. M. Dollar, Benchmarking in manipulation research: The ycb object and model set and benchmarking protocols, arXiv preprint arXiv:1502.03143.
[52] A. Farhadi, I. Endres, D. Hoiem, D. Forsyth, Describing objects by their attributes, in: The IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1778–1785.
[53] D. N. Osherson, J. Stern, O. Wilkie, M. Stob, E. E. Smith, Default probability, Cognitive Science 15 (2) (1991) 251–269.
[54] S. J. Lederman, R. L. Klatzky, Hand movements: A window into haptic object recognition, Cognitive Psychology 19 (3) (1987) 342–368.
[55] Y. Xian, B. Schiele, Z. Akata, Zero-shot learning - the good, the bad and the ugly, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[56] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, T. Darrell, Deep learning for tactile understanding from visual and haptic data, in: Robotics and Automation (ICRA), 2016 IEEE International Conference on, IEEE, 2016, pp. 536–543.