2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Haptic Knowledge Transfer Between Heterogeneous Robots using Kernel Manifold Alignment

Gyan Tatiya[1], Yash Shukla[1], Michael Edegware[1] and Jivko Sinapov[1]

***Abstract*— Humans learn about object properties using multiple modes of perception. Recent advances show that robots can use non-visual sensory modalities (i.e., haptic and tactile sensory data) coupled with exploratory behaviors (i.e., grasping, lifting, pushing, dropping, etc.) for learning objects' properties such as shape, weight, material and affordances. However, non-visual sensory representations cannot be easily transferred from one robot to another, as different robots have different bodies and sensors. Therefore, each robot needs to learn its task-specific sensory models from scratch. To address this challenge, we propose a framework for knowledge transfer using kernel manifold alignment (KEMA) that enables source robots to transfer haptic knowledge about objects to a target robot. The idea behind our approach is to learn a common latent space from multiple robots' feature spaces produced by respective sensory data while interacting with objects. To test the method, we used a dataset in which 3 simulated robots interacted with 25 objects and showed that our framework speeds up haptic object recognition and allows novel object recognition.**

I. Introduction

To recognize objects and their properties, humans use a variety of non-visual sensory modalities coupled with exploratory behaviors. While robots can use vision to recognize the shape and color of an object, camera input alone cannot determine its haptic and tactile properties, such as whether it is soft or hard, or whether it is full or empty. To perceive non-visual information, a robot must interact with the object and interpret the feedback to detect the object's characteristics. Previous works have indeed shown that robots can use non-visual sensory feedback of interaction with objects such as haptic, tactile, and/or auditory senses to perform tasks, including object recognition, object category acquisition, and language grounding (see [1], [2] for a review).

A major challenge when learning non-visual object representations is that each robot requires excessive time to perform the necessary object exploration for data collection, which prohibits rapid learning and makes it difficult to deploy non-visual object representations in practice. There is no general purpose sensory knowledge representations for non-visual features as different robots have different embodiments and sensors. As a result, it is not easy to transfer knowledge of non-visual object properties from one robot to another, so each individual robot needs to learn its task-specific sensory models from scratch.

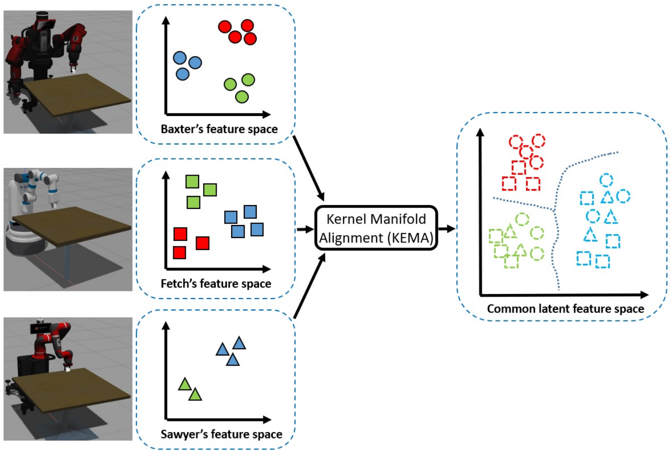

Fig. 1. Overview of the proposed framework. Feature space of different robots depict datapoints collected during object interaction. Each shape represents a robot and each color represents an object. Once each datapoint is projected into a common latent space, the decision function for a classifier is grounded in the latent space rather than the robot's own feature space.

To address this challenge, we propose a framework for haptic knowledge transfer, shown in Fig. 1, using kernel manifold alignment (KEMA) for sharing knowledge between multiple, heterogeneous robots. Our method projects the sensorimotor features of object interaction from multiple robots into a common latent space and use this latent space to train the recognition models to solve various tasks, as opposed to using each robots own sensorimotor feature space. To test our method, we collected a dataset of 3 simulated robots that performed 4 behaviors on 25 objects, and we used this dataset to transfer knowledge from two source robots to a target robot for training the target robot with less examples. The results of our experiments show that robots can bootstrap their haptic object perception skills by leveraging experience from other robots in a way that speeds up learning and allows the target robot to recognize novel objects that it has not interacted with before test time.

II. Related Work

Research in psychology and cognitive science has highlighted the significance of multiple sensory modalities used by humans to recognize objects [3], [4] and interact with them in order to learn their haptic and tactile properties [5]. Traditionally, object recognition approaches are based solely on the visual modality. More recently, several lines of recent research have proposed integrating exploratory actions with haptic modality, which has also been shown useful for learning object categories [6], [7], [8], [9], [10], [11], object

[1] Department of Computer Science, Tufts University, Email: {Gyan.Tatiya, Yash.Shukla, Michael.Edegware, Jivko.Sinapov} @ tufts.edu

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

relations [12], [13], and grounding language used to describe objects [14], [15], [16]. A remaining challenge is that non-visual sensory representations cannot be easily transferred from one robot to another, as each robot has a unique embodiment in terms of its morphology and sensor suite. As a result, each robot must interact with objects to learn its models from scratch. This work presents a knowledge-transfer framework for multiple robots that enables them to not only recognize objects with less interactions, but also to recognize novel objects without exploratory training.

To transfer knowledge, Tatiya *et al.* [17] proposed using encoder-decoder neural network to project sensorimotor features from a source robot's feature space to a target robot's feature space, allowing the target robot to classify novel objects into categories using the source robot's knowledge. One limitation was that the dataset used contained only a single robot, and thus they transferred knowledge between two physically identical robots across different behaviors. Furthermore, the method proposed would work only for two robots: the source and the target. To deal with these shortcomings, we propose a method that enables more than two robots of different embodiments to project their sensory features into a common latent space, such that the decision function for a given recognition task is grounded in the latent space rather than each individual robot's own feature space.

Domain adaptation is a transfer learning method that deals with shifts in the feature spaces of a source domain (training set) and a different but related target domain (test set). The main goal of such methods is to reduce the domain shift so that a machine learning classifier trained on the source domain can make better predictions about the target domain. Manifold alignment is a domain adaptation strategy that aligns datasets and projects them into a common latent space. Manifold alignment preserves the local geometry of each manifold and captures the correlations between manifolds, which allows knowledge transfer from one domain to another. The projected datapoints are comparable and can be used to train a single classifier for different domains.

We propose to use the kernel manifold alignment (KEMA) [18] for domain adaptation, which can align an arbitrary number of domains of different dimensionality without needing paired examples. KEMA [18] has been successfully applied to visual object recognition [18], facial expression recognition [18], and human action recognition [19]. However, KEMA has never been applied to the haptic data that robots can use for object recognition. We evaluated the performance of KEMA to adapt the sensory signals of multiple robots and obtain their aligned feature representations in a common latent space.

III. Learning Methodology

A. Notation and Problem Formulation

Let a robot perform a set of exploratory behaviors (e.g. *grasp*, *pick*), $\mathcal{B}$, on a set of objects, $\mathcal{O}$, while recording a non-visual sensory modality m (e.g. *effort*). Let the robot perform each behavior n times on each object. Let us consider $\mathcal{R}$ such robots' datasets with $\mathcal{B}_r$, m_r and n_r, where $r = 1, ..., R$. Each robot interacts with the same set of objects $\mathcal{O}$. During the i^{th} exploratory trial, the robot r observation feature is represented as $x_r^i \in \mathbb{R}^{D_r}, i = 1, ..., n_r$ where D_r is the dimensionality of the features space for robot r.

Our main goal is to learn a common latent feature space for all the $\mathcal{R}$ robots, such that the robots can be trained to recognize objects in that latent space, as opposed to each robot's own feature space. This will enable an individual robot to use the observation features collected by other robots to learn a recognition model and perform better than a model trained only using its own observation features. In addition, learning a common latent feature space would also enable a robot to recognize objects it has never interacted with, as long as other robots have. While learning the latent space, it is assumed that all the robots perform the same behavior and interact with the same set of objects.

B. Kernel Manifold Alignment (KEMA)

KEMA [18] extended the work of Wang *et al.* [20] by kernelization of the original data by transforming it into a high dimensional Hilbert space $\mathcal{H}$ with the mapping function $\phi(.) : x \mapsto \phi(x) \in \mathcal{H}$ to ensure that the transformed data is linearly separable. Due to the high dimensional feature space, the computational load would increase significantly and thus, kernel trick is used in which the problem is expressed in terms of dot products within $\mathcal{H}$. A Kernel function $K_{ij} = K(x_i, x_j) =< \phi(x_i), \phi(x_j) >$ is used to compute the kernel matrix that encodes the similarity between training examples using pair-wise inner products between mapped examples without computing $\phi(.)$ directly. We adopted Radial Basis Function (RBF) kernel as the kernel function. As there are multiple robots, R different robots' datasets are mapped into R different Hilbert spaces of dimension $\mathcal{H}_r, \phi_r(.) : x \mapsto \phi_r(x) \in \mathcal{H}_r, r = 1, ..., R$.

KEMA constructs a set of domain-specific projection functions, $\mathcal{F} = [f_1, f_2, ...f_R]^T$ that project data from R robots into a common latent space such that the examples of a same object class would locate closer while examples of different object classes would locate distantly. To achieve this, KEMA finds the data projection matrix $\mathcal{F}$ that minimizes the following cost function:

$$\begin{aligned}\{f_1, f_2, ...f_R\} &= \underset{f_1, f_2, ...f_R}{\arg\min} \left(C(f_1, f_2, ...f_R)\right) \\ &= \underset{f_1, f_2, ...f_R}{\arg\min} \left(\frac{\mu GEO + (1-\mu)SIM}{DIS}\right)\end{aligned} \tag{1}$$

where geometry (GEO) and class similarity (SIM) terms are minimized and class dissimilarity (DIS) term is maximized. The parameter $\mu \in [0, 1]$ controls the contribution of the geometry and the similarity terms. The three terms are explained as follows:

1. **Geometry (GEO)** is a matrix that represents the geometry of a domain. GEO is minimized to preserve the local geometry of each domain by penalizing projections in the input domain that are far from each other:

$$
\begin{aligned}
GEO &= \sum_{r=1}^{R} \sum_{i,j=1}^{n_r} W_g^r(i,j) \left\| f_r^T \phi_r(x_r^i) - f_r^T \phi_r(x_r^j) \right\|^2 \\
&= tr(F^T \Phi L_g \Phi^T F)
\end{aligned}
\qquad (2)
$$

where W_g^r in a similarity matrix representing the similarity between x_r^i and x_r^j, which is typically computed by k-nearest neighbor graph (k-NNG). $L_g \in \mathbb{R}^{(\sum_r n_r)\times(\sum_r n_r)}$ is a graph Laplacian matrix computed by $L_g = D_g - W_g$, where D_g is a diagonal matrix with entries $D_g(i,i) = \sum_j W_g(i,j)$.

2. **Similarity (SIM)** is a matrix that represents the class similarity of a domain. SIM is minimized to encourage examples with the same object class to be located close to each other in the latent space by penalizing projections of the same object class far from each other:

$$
\begin{aligned}
SIM &= \sum_{r,r'=1}^{R} \sum_{i,j=1}^{n_r, n_{r'}} W_s^{r,r'}(i,j) \left\| f_r^T \phi_r(x_r^i) - f_{r'}^T \phi_{r'}(x_{r'}^j) \right\|^2 \\
&= tr(F^T \Phi L_s \Phi^T F)
\end{aligned}
\qquad (3)
$$

where $W_s^{r,r'}$ in a similarity matrix that has components set to 1 if the two examples from robots r and r' belong to the same object class, and 0 otherwise. The graph Laplacian matrix is computed by $L_s = D_s - W_s$, where D_s is a diagonal matrix with entries $D_s(i,i) = \sum_j W_s(i,j)$.

3. **Dissimilarity (DIS)** is a matrix that represents the class dissimilarity of a domain. DIS is minimized to encourage examples with different object classes to be located far apart in the latent space by penalizing projections of the different object class that are close to each other:

$$
\begin{aligned}
DIS &= \sum_{r,r'=1}^{R} \sum_{i,j=1}^{n_r, n_{r'}} W_d^{r,r'}(i,j) \left\| f_r^T \phi_r(x_r^i) - f_{r'}^T \phi_{r'}(x_{r'}^j) \right\|^2 \\
&= tr(F^T \Phi L_d \Phi^T F)
\end{aligned}
\qquad (4)
$$

where $W_d^{r,r'}$ in a dissimilarity matrix that has components set to 1 if the two examples from robots r and r' belong to different objects, and 0 otherwise. The graph Laplacian is computed by $L_d = D_d - W_d$, where D_d is a diagonal matrix with entries $D_d(i,i) = \sum_j W_d(i,j)$. By combining Eqs. (2), (3), and (4), the optimization problem can be formulated as:

$$
\underset{f_1, f_2, \dots f_R}{\arg\min}\, tr\left(\frac{F^T \Phi(\mu L_g + (1-\mu)L_s)\Phi^T F}{F^T \Phi L_d \Phi^T F} \right) \qquad (5)
$$

The latent features that minimize the cost function $C(f_1, f_2, ...f_R)$ are given by the eigenvectors corresponding to the last eigenvalues of the generalized eigenproblem derived from Eq. (5) [20]:

$$
\Phi(\mu L_g + (1-\mu)L_s)\Phi^T F = \lambda \Phi L_d \Phi^T F \qquad (6)
$$

where Φ is a block diagonal matrix containing the datasets $\Phi_r = [\phi_r(x_1), ..., \phi_r(x_{n_r})]^T$, F contains the eigenvectors organized in rows for the particular domain defined in Hilbert space $\mathcal{H}_r$, where $\mathcal{F} = [f_1, f_2, ...f_H]^T$, $H = \sum_{r=1}^{R} H_r$, and λ is the eigenvalues of the generalized eigenproblem. F is in a high dimensional space that might be costly to compute. Thus, the eigenvectors are expressed as a linear combination of mapped examples using the Riesz representation theorems [21] as $f_r = \Phi_r \alpha_r$ (or $F = \Phi$ in matrix notation). By multiplying both sides by Φ^T in Eq. (6) and replacing the dot products with the corresponding kernel matrices, $K_r = \Phi_r^T \Phi_r$, the final problem is formalized as:

$$
K(\mu L_g + (1-\mu)L_s)K\Lambda = \lambda K L_d K \Lambda \qquad (7)
$$

where K contains kernel matrices K_r in a block diagonal form. The projection matrix Λ can be expressed in a block structure of size $n \times n$:

$$
\Lambda = \begin{bmatrix} \boldsymbol{\alpha_1} \\ \vdots \\ \alpha_R \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha_{1,1}} & \boldsymbol{\cdots} & \boldsymbol{\alpha_{1,n}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha_{n_1,1}} & \boldsymbol{\cdots} & \boldsymbol{\alpha_{n_1,n}} \\ \alpha_{n_1+1,1} & \cdots & \alpha_{n_1+1,n} \\ \vdots & \ddots & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,n} \end{bmatrix} \qquad (8)
$$

where the eigenvectors are highlighted in bold for the first domain, and $n = \sum_r n_r$ is the total number of examples in the kernel matrices. A new test example x_r^i can be projected to the new latent space by first mapping it to its corresponding kernel form K_r^i and then applying the corresponding projection vector α_r formulated as:

$$
P(x_r^i) = f_r^T \Phi_r^i = \alpha_r^T \Phi_r^T \Phi_r^i = \alpha_r^T K_r^i \qquad (9)
$$

where K_r^i is a kernel evaluations vector between example x_r^i and all examples of rth robot used to compute the projections α_r. For more details on KEMA, readers can refer [18], [20].

C. Object Recognition Model using Latent Features

Once the data is transferred to the latent space from multiple robots, we used the transferred data on the latent manifold to train a multi-class Support Vector Machine (SVM) [22] model with the RBF kernel to recognize different object classes. We trained two types of models: speeding up object recognition model and novel object recognition model.

To build the manifold alignment for the speeding up object recognition model, we used two source robots that are assumed to have explored the objects extensively and one target robot that is assumed to have relatively less experience with objects. To train this model, we used the transferred data from all the robots, but incrementally varied the number of examples per object used for the target robot. To test this model, we used the examples of the target robot that were not used to build the manifold alignment.

To build the manifold alignment for the novel object recognition model, we used two source robots that are assumed to have explored all the objects and one target robot that is assumed to have never explored a few objects. To train this model, we used the transferred data from two source

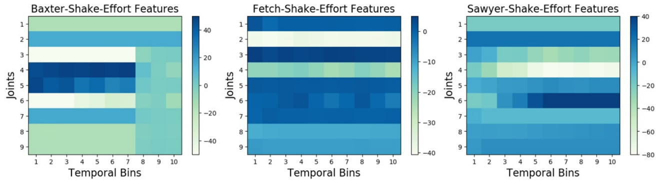

Fig. 2. Examples of *effort* features using *shake* behavior performed on an 0.62 kg block object by *Baxter*, *Fetch*, and *Sawyer* (right to left).

robots of the objects that the target robot never explored. To test this model, we used the examples of the objects that are novel to the target robot.

IV. Evaluation

A. Data Collection and Feature Extraction

A dataset was collected in which 3 simulated robots (*Baxter*, *Fetch* and *Sawyer*) perform 4 behaviors (*grasp*, *pick*, *shake* and *place*) on 25 block objects (each vary by weight from 0.01 kg to 1.5 kg). The behaviors of each robot were encoded as joint-space trajectories where the joint values are randomly sampled within a specified range of joint values for each joint of the robot. Thus, each interaction of the robot is expected to be different, which is what we would expect in the real world. During each behavior the robots recorded *effort* feedback from all joints [1]. Each behavior was performed 100 times on each object, resulting in a total of 10,000 examples (4 behaviors x 25 objects x 100 trials) per robot. Effort data was discretized into 10 temporal bins, where each bin consists of mean of effort values in that bin. Fig. 2 visualizes examples of effort features of all the robots.

B. Evaluation

To evaluate the performance of manifold alignment for knowledge transfer, we considered two tasks. In the first task, the target robot has less interaction with objects, and in the second task, the target robot has never interacted with a few objects. In both tasks, we assume both source robots have explored all the objects extensively.[2]

1) Speeding up object recognition: In this task, the main goal is to improve the object recognition performance of the less experienced target robot, by aligning the data from all the 3 robots, and then using this aligned data to train the target robot. For the baseline condition, the target robot is trained to recognize objects by using its own data collected during object interactions. For the transfer condition, the target robot is trained to recognize objects by using the aligned data in the latent feature space corresponding to all the 3 robots. We incremented the number of examples per object used to train the target robot from 1 to 80, and we used the held-out 20 examples for testing. For both conditions, we performed 5-fold cross validation such that each example is included in test set once and computed accuracy $A = \frac{\text{correct predictions}}{\text{total predictions}}\%$, and reported average accuracy of all the folds.

2) Novel object recognition: In this task, the goal is to enable the target robot to recognize n objects it never interacted with. Both source robots interact with all the 25 objects, while the target robot interacts with only $25 - n$ randomly selected objects. The $25-n$ objects shared by all 3 robots are used to build the manifold alignment that transfers the sensory signal of the robots to the latent space. Then a classifier is trained using the transferred data of the source robot corresponding to the objects that are novel to the target robot. Subsequently, to test this classifier, the transferred data of the n objects that the target robot did not interact with is used that were not used to build the alignment. Similar to speeding up object recognition, we reported the accuracy of this classifier to evaluate its performance and compared it with the chance accuracy of the classifier. The process of selecting $25 - n$ objects randomly to build the manifold alignment, training the classifier using transferred data of the source robots and testing the classifier on n novel objects was repeated 10 times to produce an accuracy estimate.

V. Results

A. Illustrative Example

Consider the case where the 3 robots perform the *place* behavior on all 25 objects 10 different times while recording *effort* signals, which were used to build the manifold alignment using KEMA and generate latent features. We plotted the first two dimensions of the latent features, and reduced the dimensionality of the original sensory signal to 2 by Principal Component Analysis. As shown in Fig. 3, the datapoints collected by the 3 robots of 5 different objects are clustered together in the common latent space.

B. Speeding up object recognition results

Fig. 4 shows the object recognition performance, where *Baxter* and *Sawyer* serve as the source robots and *Fetch* serves as the target robot. To build the manifold alignment, we incrementally varied the number of interactions of the target robot from 1 to 80, and to test the classifier, held-out 20 examples are used. Note that to choose the amount of source robot data for building alignment and number of dimensions of latent features used to train the model, we performed a grid search, in which we experimented with different amount of source robot data and different number of dimensions and used the optimal parameters for the final results. Generally, if the target robot interacts less with objects, using more source robots' data generates better latent features, and using the first 1 or 2 dimensions of the latent features achieves high accuracy as they are the most correlated dimensions among all the robots.[3] Fig. 4 compares the recognition accuracy of the baseline condition, where the target robot learns to recognize objects using only its own features, and the transfer condition, where the target robot learns to recognize objects

[1] The sampling rate of *Baxter* is 50Hz, and *Fetch* and *Sawyer* is 100Hz. All the robot's arm have 9 joints including 2 grippers.

[2] Datasets, source code and complete results for study replication are available at: `https://github.com/gtatiya/Haptic-Knowledge-Transfer-KEMA`.

[3] Note that using entire source robots' data and latent features for training the target robot did not perform better than using optimal amount of source robot data and number of latent features.

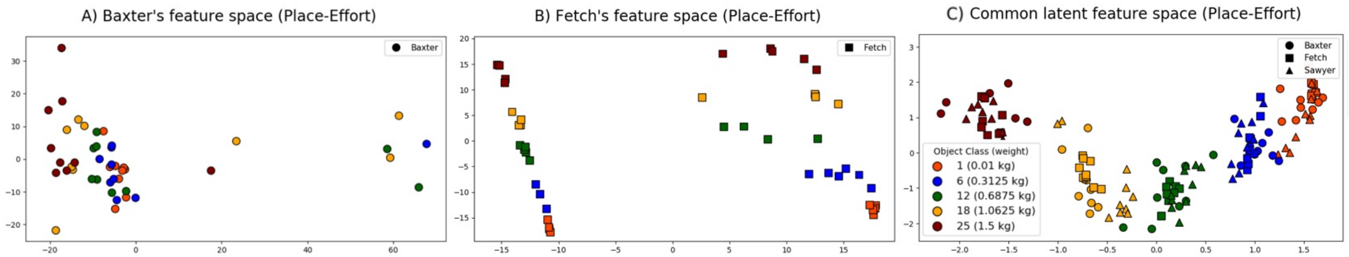

Fig. 3. Original sensory features of (A) Baxter and (B) Fetch for *place-effort* performed on 5 objects in 2D space, and first 2 dimensions of corresponding features in the common latent feature space (C).

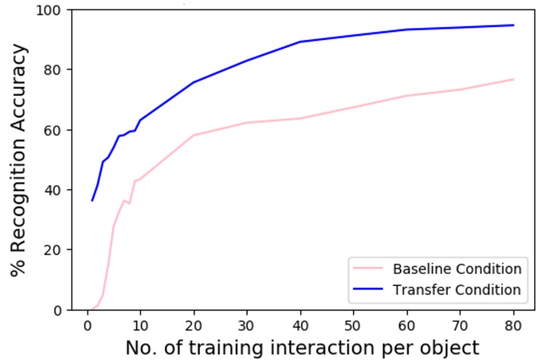

Fig. 4. Accuracy of the baseline and transfer conditions, where *Fetch* serves as the target robot, and *Baxter* and *Sawyer* serve as the source robots.

using its own as well as the source robots' latent features. In both conditions, the recognition accuracy is computed by performing a weighted combination of all the behaviors based on their performance on the training examples.

For most behaviors, the transfer condition performs consistently better than the baseline condition. A significant boost in performance is observed with a fewer number of the target robot's interactions per object. Fig. 4 shows that by performing all the behaviors with each object only once, the target robot achieves around 0% accuracy in the baseline condition, whereas it achieves 36.28% accuracy in the transfer condition. This result indicates that in cases where the target robot has limited time to learn the task, transferring knowledge from other robots can speed up as well as improve the classification performance. We also experimented with *Baxter* and *Sawyer* as the target robot, and the other 2 robots as the source robot, and observed similar boost in performance in the transfer condition.

C. Novel object recognition results

For a case where the *Fetch* robot has not interacted with 2 of the objects, we trained a classifier using the latent features of the source robots (*Baxter* and *Sawyer*) performing the *place* behavior on those objects. Fig. 5 visualizes the data used to train and test the classifier. In Fig. 5A, squares with blue and red outline show the source robots' training data and circles show the true labels of the target robot's data used to test the classifier. Each color represents a different object. Fig. 5B shows the predictions of the classifier, which is able to correctly classify 100% of the test data.

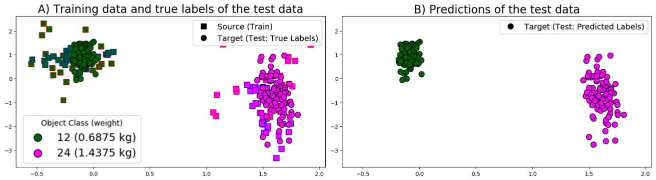

Fig. 5. Visualization of the training and testing datapoint used to train the target robot (*Fetch*) to detect 2 novel objects in 2D space. (A) shows the training data in squares corresponding to the source robots (*Baxter* and *Sawyer*) latent features of *place* behavior, and the test data in circles corresponds to the true labels of the target robot (*Fetch*). (B) shows the predictions of the test data, which is 100% correct.

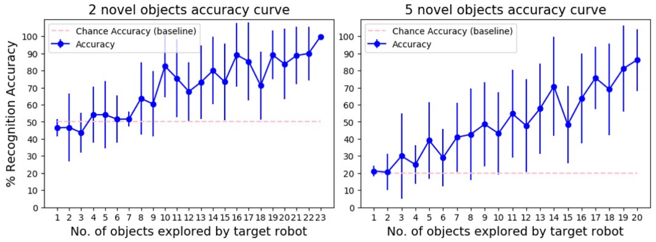

Fig. 6. Accuracy curve of the target robot (*Fetch*) for detecting 2 and 5 novel objects (left to right) for different number of objects explored by it using the knowledge transferred by the source robots (*Baxter* and *Sawyer*).

Fig. 6 shows the results when the target robot (*Fetch*) was trained to recognize 2 and 5 novel objects by incrementing the number of objects explored by the target robot used to build the manifold alignment. To build the manifold alignment, 30% of the source robots' data (*Baxter* and *Sawyer*) was used. In most cases, the target robot achieves better than chance accuracy, and as the target robot interacts with more objects, its performance to recognize novel objects improves. Thus, the target robot can learn to recognize objects it never interacted with by using the knowledge transferred by the source robots. Similar results were observed when the *Baxter* and *Fetch* serve as the target robot.

D. Heterogeneous Feature Representation

A robot's sensory features can be represented in different ways depending on the feature extraction method. To evaluate our framework with different feature representations used by the individual robots, we discretized the effort data into 15 temporal bins, where each bin consists of effort values' range computed by subtracting the minimum effort value from the maximum effort value in that bin. Fig. 7 shows

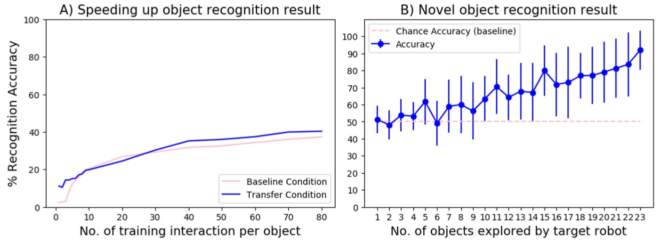

Fig. 7. Results of a different feature representation, where *Baxter* and *Sawyer* serve as the source robots and *Fetch* serves as the target robot. (A) shows the results of the speeding up object recognition task, where predictions of all the behaviors are combined. (B) shows the accuracy curve of 2 novel objects recognition task.

the results of the speeding up object recognition and the novel object recognition tasks on this new representation, where *Baxter* and *Sawyer* serve as the source robots and *Fetch* serves as the target robot. Fig. 7A indicates that the transfer condition enables the target robot to perform better than the baseline condition especially with less experience with objects. Moreover, Fig. 7B suggests that the target robot learned to recognize novel objects with knowledge transferred by the source robots. These results are consistent with the results of the previous feature representation we presented, which means knowledge can be transferred using KEMA for different representations.

VI. Conclusion and Future Work

To enable robots to work in human-inhabited environment, they would need to recognize objects' properties through interaction. Non-visual sensory signals (e.g. haptic) collected by a robot's interaction cannot be used to train another robot as the feature space of such data is different for robots with different embodiments. In addition, collecting interaction based sensory signals is a time consuming process. Thus, we propose using kernel manifold alignment, to align the feature spaces of different robots into a common feature space, and use it to train the robots. We showed that our approach can enable the target robot to not only speed up the learning process by learning with less interaction, but also perform better by using aligned features from other robots rather than learning just from its own features. Moreover, we showed that the target robot can learn to recognize novel objects by knowledge transferred by the source robots.

A limitation of our experiment is that the dataset we used contains simulated robots, thus in future work, we plan to test our proposed knowledge transfer method on real robots. A kernel function that is designed to specifically capture time series data such as haptics is also a promising avenue for future exploration. Moreover, we would adapt our knowledge transfer method to a larger variety of non-visual sensors other than effort such as audio, temperature, and vibration. Finally, in our experiments, we addressed the object recognition task. In future work, we plan to extend our method to handle sensory knowledge transfer for other tasks, such as object manipulation, and language grounding.

Acknowledgements

This work was conducted with partial support from DARPA (W911NF-19-2-0006), the Verizon Foundation, PTC Inc., and the Center for Applied Brain and Cognitive Sciences (CABCS).

References

[1] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1273–1291, 2017.

[2] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A review of tactile information: Perception and action through touch," *IEEE Transactions on Robotics*, 2020.

[3] T. Wilcox, R. Woods, C. Chapa, and S. McCurry, "Multisensory exploration and object individuation in infancy." *Dev. Psy.*, 2007.

[4] M. O. Ernst and H. H. Bülthoff, "Merging the senses into a robust percept," *Trends in cognitive sciences*, vol. 8, no. 4, pp. 162–169, 2004.

[5] E. J. Gibson, "Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge," *Annual review of psychology*, vol. 39, no. 1, pp. 1–42, 1988.

[6] J. Sinapov, C. Schenck, K. Staley, V. Sukhoy, and A. Stoytchev, "Grounding semantic categories in behavioral interactions: Experiments with 100 objects," *Robotics and Autonomous Systems*, 2014.

[7] V. Högman, M. Björkman, A. Maki, and D. Kragic, "A sensorimotor learning framework for object categorization," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 1, pp. 15–25, 2016.

[8] Z. Erickson, S. Chernova, and C. C. Kemp, "Semi-supervised haptic material recognition for robots using generative adversarial networks," in *Conference on Robot Learning*, 2017.

[9] G. Tatiya and J. Sinapov, "Deep multi-sensory object category recognition using interactive behavioral exploration," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

[10] S. Jin, H. Liu, B. Wang, and F. Sun, "Open-enviroment robotic acoustic perception for object recognition," *Frontiers in Neurorobotics*, vol. 13, p. 96, 2019.

[11] R. Braud, A. Giagkos, P. Shaw, M. Lee, and Q. Shen, "Robot multimodal object perception and recognition: Synthetic maturation of sensorimotor learning in embodied systems," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[12] J. Sinapov, P. Khante, M. Svetlik, and P. Stone, "Learning to order objects using haptic and proprioceptive exploratory behaviors." in *IJCAI*, 2016, pp. 3462–3468.

[13] T. Taunyazov, W. Sng, H. H. See, B. Lim, J. Kuan, A. F. Ansari, B. C. Tee, and H. Soh, "Event-driven visual-tactile sensing and learning for robots," *Robotics: Science and Systems*, 2020.

[14] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. M. Perez-Tejada, M. Arrigo, T. Darrell, and K. J. Kuchenbecker, "Robotic learning of haptic adjectives through physical interaction," *Robotics and Autonomous Systems*, vol. 63, pp. 279–292, 2015.

[15] J. Thomason, J. Sinapov, R. J. Mooney, and P. Stone, "Guiding exploratory behaviors for multi-modal grounding of linguistic descriptions," in *32nd AAAI Conference on Artificial Intelligence*, 2018.

[16] B. Richardson and K. Kuchenbecker, "Improving haptic adjective recognition with unsupervised feature learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.

[17] G. Tatiya, R. Hosseini, M. C. Hughes, and J. Sinapov, "Sensorimotor cross-behavior knowledge transfer for grounded category recognition," in *International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. IEEE, 2019.

[18] D. Tuia and G. Camps-Valls, "Kernel manifold alignment for domain adaptation," *PloS one*, vol. 11, no. 2, p. e0148655, 2016.

[19] Y. Liu, Z. Lu, J. Li, C. Yao, and Y. Deng, "Transferable feature representation for visible-to-infrared cross-dataset human action recognition," *Complexity*, 2018.

[20] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *Twenty-second international joint conference on artificial intelligence*, 2011.

[21] F. Riesz and B. S. Nagy, "Functional analysis, frederick ungar pub," *Co., New York*, 1955.

[22] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, 1998.