Haptic Object Recognition for Multi-Fingered Robot Hands

Stefan Escaida Navarro, Nicolas Gorges, and Heinz Wörn
Institute for Process Control and Robotics
Karlsruhe Institute of Technology
e-mail: {escaida, gorges, woern}@kit.edu

Julian Schill, Tamim Asfour, and Rüdiger Dillmann
Humanoids and Intelligence Systems Lab
Karlsruhe Institute of Technology
e-mail: {schill, asfour, dillmann}@kit.edu

ABSTRACT

In this paper, we present an approach for haptic object recognition and its evaluation on multi-fingered robot hands. The recognition approach is based on extracting key features of tactile and kinesthetic data from multiple palpations using a clustering algorithm. A multi-sensory object representation is built by fusion of tactile and kinesthetic features.

We evaluated our approach on three robot hands and compared the recognition performance using object sets consisting of daily household objects. Experimental results using the five-fingered hand of the humanoid robot ARMAR, the three-fingered Schunk Dexterous Hand 2 and a parallel Gripper are performed. The results show that the proposed approach generalizes to different robot hands. Keywords: Object Recognition, Bag of Keypoints, Robot Hands, Tactile Sensing.

1 INTRODUCTION

Several applications of artificial haptic sensors have been envisioned and studied over the past years. Especially in humanoid robotics the interest is given by the goal of reproducing human-like capabilities. A strong haptic sensing capability is essential for autonomous action in unstructured environments as well as human-robot interaction. The application domains that have been investigated for touch sensors include contact detection, grasping, dexterous manipulation, slip detection and object recognition. This work concentrates on the latter topic with the major contribution being the conceptual improvement of a previously proposed concept including the key steps necessary for using the system with different types of robot hands varying in the number of DoF and tactile sensors.

Various approaches have been proposed for haptic object recognition. In general it is difficult to compare and reproduce the results, as the approaches use different and very often unique hardware. They lack generality or have not been tested across several platforms. Some approaches consider the use of only one sensor type, that means, either tactile or kinesthetic modality [15]. A range of approaches concentrate on classifying an object based on a point cloud acquired from haptic data [2, 12]. These approaches are usually limited by the requirement of objects being fixed and stationary. Therefore, recent work tries to build an object model directly from haptic sensor data without building a 3D-model. In [10], objects were categorized according to their shape perceived through a single grasp.

A popular approach for pattern classification in vision is the so called *bag of keypoints* that has been studied extensively [5, 1]. Applied to haptics it means that, like the human, when solving the task of haptic object recognition, the robotic system relies on multiple, partial observations. This has been studied by Schneider et al. [13] and in our previous work [8]. These works report that a growing number of combined observations improves haptic object recognition giving support to the bag of keypoints idea. This approach also has the advantage of relying on the data as observed by the robot hand without reconstructing geometric properties explicitly. This way the problem of spatial registration of observations can be avoided, meaning objects do not have to be fixed in the recognition process.

In contrast to the work by other authors, our work evaluates kinesthetic and tactile information separately, giving insight into importance of tactile and kinesthetic information for the recognition process. Objects are held by the robot hands without support during data acquisition, leading to high variability in the object's position and orientation. Additionally, the evaluation on diverse robotic hands demonstrates the universality of the approach.

After this introduction, the object recognition framework is introduced in section 2 followed by section 3, which explains how the components of the recognition framework are linked. Then the evaluated robotic hands and the tactile sensor system are presented in section 4. The results of our approach are presented in section 5 and finally the conclusions are given in section 6.

2 OBJECT RECOGNITION FRAMEWORK

On one hand the algorithms presented in this section deal with the elementary steps that transform haptic data from one representation to another and on the other hand they describe how to build an object representation from these intermediate steps.

2.1 Principal Component Analysis

A principal component analysis (PCA) is used for identifying a transform from a high dimensional feature space to a lower dimensional subspace that best accounts for the variance present in the data. In our work it is used to find characteristic features of contact pattern examples (see sec. 3). For the PCA n features $\vec{v}_i$ of dimensionality s are given. The mean vector is computed to

$$\vec{\mu} = \sum_{i=1}^{n} \frac{1}{n} \vec{v}_i. \tag{1}$$

The covariance matrix is given by

$$\underline{K} = \frac{1}{n} \sum_{i=1}^{n} (\vec{v}_i - \vec{\mu})(\vec{v}_i - \vec{\mu})^T \tag{2}$$

which is of size $s \times s$.

The principal component analysis of $\underline{K}$ results in s eigenvalues Λ_i and the accordant eigenvectors γ_i, which span the orthogonal eigenspace $\Gamma = (\gamma_1, \cdots, \gamma_p)$. This eigenspace describes the highest variance between the features. The $s \times s$-matrix, with each eigenvector as a column, is reduced to a $s \times d$-matrix $\underline{E}$ by taking only the first d eigenvectors with highest eigenvalues. Given an image in vector presentation $\vec{v}$ and the reduced eigenvector matrix $\underline{E}$ the image vector is reduced to a $d \times 1$-vector by $\vec{k} = \underline{E}^T \cdot \vec{v}$ which is used as the final presentation of a feature.

If using PCA for representing tactile features a tactile image $\underline{I}$ of the size $w \times h$ can be represented as the vector $\vec{v}$ of size $s = w \cdot h$.

$$\underline{I} = \begin{pmatrix} \underline{I}_{11} & \cdots & \underline{I}_{w1} \\ \vdots & \ddots & \vdots \\ \underline{I}_{1h} & \cdots & \underline{I}_{wh} \end{pmatrix}$$

$$\Rightarrow \vec{v} = \left(\underline{I}_{11}\dots\underline{I}_{w1}, \underline{I}_{12}\dots\underline{I}_{w2}\dots\underline{I}_{1h}\dots\underline{I}_{wh}\right)^T$$

In a PCA context, the intrinsic dimensionality is the number of eigenvectors γ_i that are sufficient to represent the variance present in the data. The *dimensionality reduction toolbox* [16] that was used in this work solves this with a threshold. For thresholding, the eigenvalues are normalized to $[0,1]$ and eigenvectors with eigenvalues $< 0{,}025$ are discarded.

2.2 Normalization of Contacts

A normalization of contacts may be profitable if the variance in the tactile data does not represent variance in the object's appearence.[1] The grey values of an imprint vary depending on the applied pressure. Thus, a tactile image is simply normalized for pressure by applying histogram equalization or a binarization procedure. As the data of the tactile sensor matrix corresponds to a two-dimensional planar image, these images are additionally analyzed using moments up to the 2^{nd} order [9]. The moment analysis results into the position and the orientation amongst other features. Using this information, the centroid of the imprint is shifted into the center of the image and the orientation of the imprint is aligned to the image axes, as shown in Fig. 1.

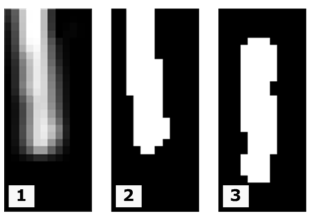


Figure 1: Steps of the normalization procedure: (1) Starting from a raw tactile image. (2) The image is normalized for pressure. (3) The imprint is moved to the image center and the axes of the imprint are aligned to the image axes.

2.3 Self-Organizing Maps

A clustering algorithm is needed for describing a feature space in terms of salient features. A self-organizing map (SOM) [11] is a neural network based clustering algorithm that describes a structure-preservative mapping from high-dimensional input space to a 2D map. Similar patterns in the input space lie in a geographical near position on the 2D map.

The SOM is described by a set of neurons c_i. Each c_i is assigned to a n-dimensional weight vector m_i and a position r_i on the map. Each neuron represents a currently unknown category/cluster. During the training process each input pattern is assigned to one of these categories and simultaneously the discrimination between the categories becomes more precise with each step of the training. The SOM can be referred to as an unsupervised learning or clustering method.

One training run t consists of 4 steps:

1. Pick a random input vector $x(t)$.

2. Calculate distance between weight vectors and input vector:

$$D_i(t) = \|x(t) - m_i(t)\|.$$

3. Determine winner neuron c:

$$c = \operatorname*{argmin}_i(D_i(t)).$$

4. Adapt weight neurons in the neighborhood of the winner:

$$m_i(t+1) = m_i(t) + \alpha(t)\cdot h_{ci}(t)\cdot[x(t) - m_i(t)].$$

The neighborhood function is given by:

$$h_{ci}(t) = \exp\frac{\|r_c - r_i\|^2}{2\cdot\sigma(t)^2}.$$

An adaptation means the movement of the weight vectors into the direction of the input vectors. The learning process terminates by reducing the learning rate $\alpha(t)$ and the neighborhood $h_{ci}(t)$. The SOM converges to a stable state if no further changes occur.

The use of a $w\times h$ SOM results into $w\cdot h$ categories for describing an input pattern. After training, the SOM can be used for assigning a pattern to one of the determined classes by the function $S(x)$:

$$p = S(x) = \operatorname*{argmin}_i(\|x - m_i\|) \in [1, w\cdot h] \tag{3}$$

In a vector representation this can be written as a $w\cdot h$ zero vector with a value one at position p.

2.3.1 Activation

A soft decision can be made by weighting each weight neuron by its distance to the input pattern. For a given feature vector x an *activation vector a* of length $w\cdot h$ is calculated as follows: If x is closest to the vectors m_i, then the activation value $a_i < 1$ is related to the respective distances from x to every m_i. Figure 2 shows an example of a SOM and a activation vector calculated using the distances between the clusters (neurons).

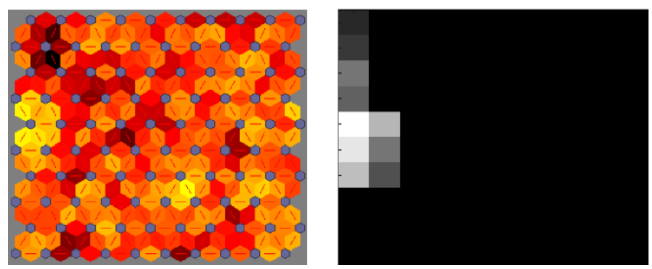


Figure 2: Left: distance relationship of the neurons (indicated by blue points) on a SOM. Bright colors correspond to lower distances, dark colors higher to distances. Right: 2D representation of an activation vector.

2.4 Bag of Keypoints Approach

For the bag of keypoints approach a set of salient features is determined by unsupervised learning/clustering on a representative dataset of the object features. In our case such a training set X_T consists of tactile imprints or joint configurations. From the clustering algorithm with input X_T and an integer k results the keypoint set $X_d = \{x_1,\dots,x_k\}$.

A vector $d = \{d_1,\dots,d_k\}$ we call *descriptor* is used for representing an object. The descriptor is built up by summing the activation vectors a from 1 or more observations; d is normalized at the end to a sum value 1 in a fashion similar to histograms. This approach can cope with partial observations and still result in a model for the complete object if dense data is available. Even more so, descriptors based on few observations and descriptors based on many observations of the same object can still be matched against each other.

We use the activation vectors described in sec. 2.3.1 rather than matching observations against a single keypoint, as a strategy for

[1] In our experiments this was only the case for the Gripper (see sec. 5).

coping with big numbers k of clusters, needed for accurately describing the feature space when using SOMs. When comparing descriptors of objects, we want to ensure similar observations are matched even when the keypoints for each palpation don't match exactly. Therefore the activation vectors can be seen as a way of capitalizing on the local similarity structures of the SOM.

2.5 Artificial Neural Network Classifier

An artificial neural network (ANN) is described by a set of neurons n_i and a set of edges e_i which connect the neurons and build a network. A single neuron has several inputs and one output. Each neuron processes the input with a nonlinear weighted sum and generates the output, which is then propagated to the connected neurons. Each neuron n_i is assigned to exactly one layer. Usually a network of neurons consists of an input layer, a hidden layer and an output layer. The size of the input layer refers to the size of feature size, whereas the size of the output layer is based on the number of classes. A feed-forward network allows the signals to travel only one way from the input layer to the output layer. During the training process, the weights and the thresholds of the neurons are adapted so that the error with respect to the target outputs is minimized. An artificial neural network can be used as a classifier. Given a feature as input to the network, the output neuron with the highest activation represents the assigned class.

3 SYSTEM OVERVIEW

The single components presented in the previous section converge to a haptic object recognition system as illustrated on fig. 3. The flow of the black arrows follows the processing steps from the raw data of K palpations to the object representation used for classification. Green boxes correspond to the representation haptic data has at a given intermediate step. Blue boxes represent the machine learning algorithms used to process the haptic data; their parameters are estimated in the training phase which is indicated by the red arrows and boxes with red contours. The learning of parameters is done incrementally: after one parameter set is estimated, the training data is propagated to the next stage in the system to be used again. In the training phase all available data is presented to the machine learning algorithms at once and labeled features are only used for training the ANN-Classifier. For the evaluation phase unseen palpations are processed in the same steps and confronted with the ANN-classifiers.

In the system a late fusion of the tactile and kinesthetic modalities is favored over the early fusion, because they belong to different feature spaces. The advantage of this approach is that it is possible to examine the representational power of tactile and kinesthetic data independently and at the same time estimate the boost in recognition provided by the fusion. While the kinesthetic feature extraction is very straightforward, i.e. data from joint encoders are fed directly to the SOMs, the imprints from the tactile sensors are treated separately at first. Because of the high dimensionality of the imprints a compact representation is found with PCA. Alternatively extracting describing the imprints by the contact's center-point and shape (elliptical features)[9] did not perform better in our experiments.

At the heart of our approach lies the bag of keypoints approach, described in sec. 2.4. It admits any number K of palpations whose features where previously extracted to be considered for a descriptor (object representation). Also, it is this step that embodies the flexibility needed for the use of different robot hands. For each of the N tactile sensor a dedicated descriptor is built; then the N descriptors are concatenated resulting in a descriptor for the tactile modality. A global object representation is obtained by concatenating the descriptors for both modalities. This procedure is independent of the degrees of freedom and number of tactile sensors present in the hand.

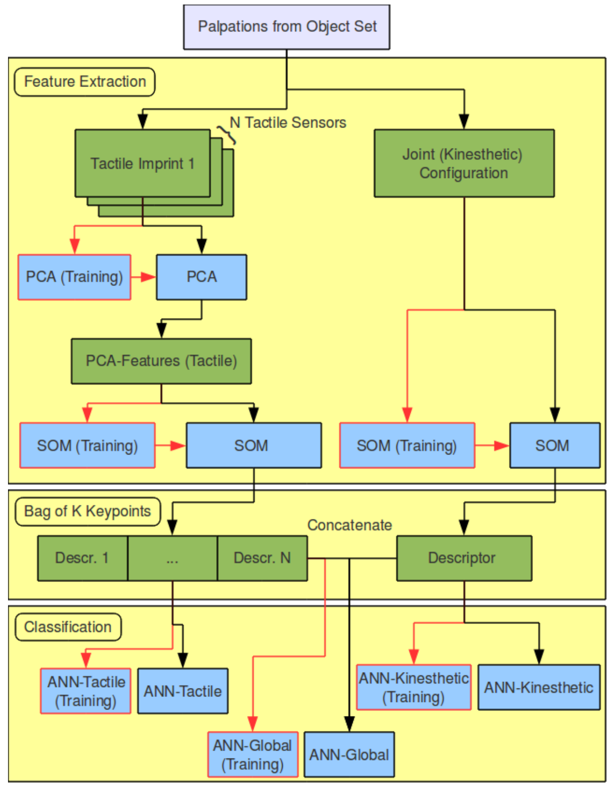


Figure 3: Flow of haptic data from perception to recognition.

3.1 Classifier Choice

In the previous work [8] we used a Bayesian classifier and modeled the object classes in descriptor feature space as Gaussian distributions. The decision to use a ANN-Classifier in this work is motivated by an analysis of the feature space, again using the SOMs. Johnsson and Balkenius [10] used a similar approach, where SOMs are used in a final step to infer the shape of objects. In our analysis, first a clustering with SOMs is performed on the descriptor features obtained in the training phase. Then, for each object class the descriptors are mapped onto the respective winner neurons of the SOM. By labeling the results it is possible to visually identify the regions in the SOM corresponding to a certain object class. The result of one analysis with the object set from our experiments (see fig. 8(b)) is displayed in fig. 4.

The descriptors of object classes in fig. 4 gather in one or more clusters and therefore we infer that in general a single Gaussian distribution is not enough to model the object classes correctly. Also, some regions in the SOM show overlapping which can be attributed to the fact, that there the SOM training algorithm does not use a weighted distance metric. The ANN-Classifier can cope with multimodal distributions and find appropriate weights for the dimensions in feature space thus being a good choice for this problem.

4 ROBOT EVALUATION PLATTFORMS

We evaluated the system on three different robot hand platforms, including a parallel Gripper, the Schunk Dexterous Hand 2 and the anthropomorphic robot hand of the humanoid Robot ARMAR-IIIb. All of the robot hands use the same tactile sensor technology by Weiss Robotics.

4.1 Tactile Sensors

The tactile sensor systems of the evaluated robot robotic hands consist of tactile sensor pads by Weiss Robotics [18]. The sensors use

(a) WRT102 Parallel Gripper (b) Schunk Dexterous Hand 2 (c) Anthropomorphic ARMAR-IIIb Hand

Figure 5: The robotic hands used for evaluation. They differ in their mechanical design, the degrees of freedom and the number of tactile sensor patches.

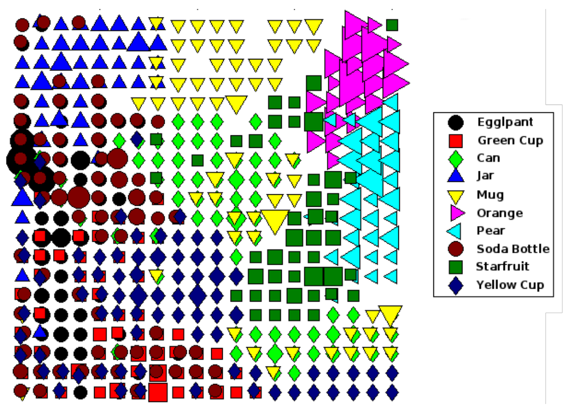


Figure 4: Self-Organizing-Map of size 20×20 helps visualizing the descriptor feature space structure. The markers on the neurons indicate one or more descriptors were mapped; the size of the markers relates to the number hits.

a resistive working principle to pick up a pressure profile. This is realized by an array of electrodes which is covered with conductive foam. A decrease of electrical resistance can be measured when a pressure is applied to the foam. Further details on the working principle of the sensors can be found in [6, 7, 17].

4.2 Parallel Gripper

The WRT102 parallel Gripper is a simple manipulator with one degree of freedom (DoF) (see fig. 5(a)). It consists of the SCHUNK PG-70 Gripper equipped with two tactile sensing pads DSA 9205 from Weiss Robotics, each with 7×14 taxels. The resolution is higher compared to the ones used for the anthropomorphic hand which enables the Gripper to take bigger imprints of the objects. The Gripper has less tactile sensor patches and less DoFs than the other robot hands, but the accuracy of the Gripper's positioning sensors is better and the resulting signal directly refers to the partial size of a grasped object. The WRT102 Gripper can determine its position with a precision superior to 1 mm. A real drawback of the Gripper is the maximal jaw width of 70mm. This limits the object candidates to be grasped to rather small objects.

4.3 Anthropomorphic ARMAR-IIIb Hand

The humanoid robot ARMAR-IIIb [3] is equipped with a FRH-4 anthropomorphic hand, which is described in [4] and is shown in Fig. 5(c). The hand is driven pneumatically using fluidic actuators and it has eleven joints: Two for each finger and one for the palm. The actuators of the ring and pinkie finger have been coupled, so that the hand has 8 independent controllable DoFs. Modifications to the original design have been made. Joint encoders and pressure sensors were attached for force-position control. Additionally tactile sensors have been mounted on the fingertips and in the palm of the hand. For this work six tactile sensor patches are used: Three in the palm with a resolution of 4×6 and one on each finger tip of the thumb, index finger and middle finger with a resolution of 4×7.

4.4 Schunk Dexterous Hand 2

The Schunk Dexterous Hand 2 (SDH-2) [14] is a electrically driven three finger hand and is shown in Fig. 5(b). It has 7 DoFs: two in every finger and two fingers can change their position allowing many different grasps. The gripping force, speed and position can be flexibly controlled with the integrated control electronics. The hand is equipped with tactile sensors in the finger tips and the phalanges. The sensor arrays in the finger tips have 13×6 taxels, while the resolution of the sensor pads in the phalanges is 14×6.

5 EXPERIMENTS

Following the bag of keypoints idea, in our scenario objects are grasped repeatedly for recognition. The position and orientation of an object in the hands is only constrained by the effect of an adaptive grasp. Objects are not supported by a planar surface and are therefore accessible from every side. Not taken into account for the recognition experiments are configurations of objects in which the grasping is not stable. In the experiments objects are passed on to the robotic hands by the human and recognition is evaluated after a small series of grasps.

The experiments were conducted on two sets of common household objects. The Gripper can only afford small objects and therefore has it's own object set of 13 objects. The other robot hands are evaluated on sets with 10 objects; 7 of them are the same for both whereas 3 of them where not firmly graspable for both hands.

Unless noted otherwise parameters were shared across the different experiments. Evaluation was performed in a 3-fold cross validation setting with 45 samples per object altogether. 30 samples are used for training and 15 samples for evaluation each time. The size of tactile and kinesthetic SOMs was chosen 10×10. The target dimension of the PCA transform for tactile imprints is 9 as a result from estimating intrinsic dimensionality (it was the same for all hands). Every ANN-Classifier used was trained with one hidden

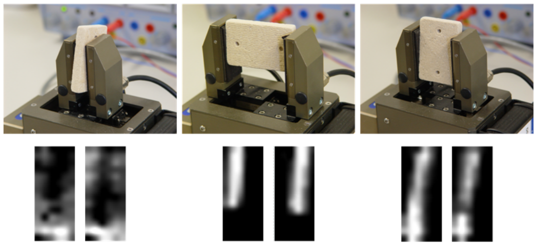


Figure 6: Three tactile imprints obtained while grasping a rectangular cuboid.

layer consisting 40 of neurons, once for the global classifier and for each modality.

Descriptors representing an object are built by combining 4 palpations. Reusing the samples in different combinations leads to new descriptors which is profitable for the training and evaluation. To avoid combinatorial explotion care is taken so that no subset of 3 palpation samples can participate in two descriptors at once, in other words, a maximum of 2 samples are shared across each pair of descriptors. As a result the samples are uniformly distributed in the descriptors and finally the system is trained with 945 descriptors and evaluated with 105 descriptors per object in each cross validation iteration.

In the following subsections we give more details about the experimental setup for each hand.

5.1 Gripper Setup

In the experiments the Gripper lies on a planar surface. The objects of fig. 8(a) are held in a random orientation between the brackets and the Gripper is closed, coming to a halt once it cannot move further. A palpation is registered when this state is stable. Fig. 5.1 shows an example of this procedure and the resulting tactile imprints.

To avoid unwanted variance brought into the data by the random orientation of the objects, tactile patterns are normalized with respect to position and orientation according to sec. 2.2. This is important, because this step is not performed for the hands, where information about the contact position and orientation is relevant. Also, this constitutes a difference to the work of Schneider et al. [13] where the orientation of contacts is relevant, because objects are standing on a planar surface when being explored.

5.2 Schunk Dexterous Hand 2 Setup

For the SDH-2 experiments the hand was also mounted on a planar surface with the palm pointing upwards. The 2 opposing fingers were in a parallel configuration, the objects were held randomly in between the fingers and a reactive grasp procedure was initiated. First the phalanges of the fingers close until the tactile sensors on the fingers show a contact. Then the fingertips are closed leading to a enveloping grasp. Examples for some of the grasps can be seen in fig. 5.2.

Figure 7: Four different grips of the eggplant (plastic) with the SDH-2 hand.

Table 1: Recognition Rates (%)

Manipulator	Objects	RR Total	RR Kin.	RR Tact.
Gripper	13	92.25	88.54	50.37
SDH-2	10	93.30	77.07	78.88
ARMAR-IIIb	10	90.06	81.96	68.63

5.3 ARMAR-IIIb Hand Setup

For the experiment on ARMAR-IIIb the objects of fig. 8(c) are slightly pushed against palm of the Hand. The robot detects this with the mounted 6-DOF force-torque sensor and closes the palm joint, index finger, middle finger, and thumb. The ring finger and pinkie are not used, as there are no active tactile sensors and the joints are coupled, giving only inaccurate sensor data. Because of the compliance of the fluidic actuators the grasped object is enveloped by the closed three fingers of the hand with a force grip. The data of the tactile sensors and joint encoders is then collected and used for the methods described above.

5.4 Results

Table 1 gives an overview of the recognition rates for the experiments, including the recognition rates for single kinesthetic and tactile modalities. In every experiment our system achieves good recognition rates and the numbers also illustrates that the fusion of kinesthetic and tactile modalities rewarding. The confusion matrices in fig. 9 provides some more details for single objects.

In the Gripper experiment the kinesthetic modality dominates the tactile modality in single mode recognition performance. The increase in the global recognition with respect to the kinesthetic is comparatively small. Firstly, this can be explained by the high positioning sensing capabilities of the Gripper – it can be compared to measuring the objects with a ruler. Secondly, the tactile data is reduced to appearance, because rotation and position of contacts is of no use in this case. The spoon (11) and the screw driver (9) have the lowest recognition performance of the set. Interestingly, they get confused with each other reciprocally most likely due to the handles that are similar.

For both of the hands we can observe a small amount of confusion between the pear (9) and the orange (2). In fact these objects have a very similar size and again we observe that confusion tends to be mutual. The same can be said for the salt (6) and the can (5) from the ARMAR-IIIb hand. Perhaps the most interesting observation is that on these robotic hands recognition on tactile modality has a bigger contribution. Probably this can be explained by two factors: firstly, measurement with the joint encoders on the hands is not as precise as the Gripper and secondly, on the hands more distributed tactile sensing is available. Tactile sensors distributed on the fingers and the palm (ARMAR-IIIb Hand) constitutes an enveloping sensing system. Because touch is required for sensing, contact information not only accounts for local shape information but for global shape information also.

6 CONCLUSIONS

In this work we showed an improved haptic object recognition framework that is flexible and performs well on different robotic hands. The comparison of the results obtained with each manipulator gives insight into the use of tactile sensors and joint encoders for robot hands.

In the future we will explore the possibilities of linking our approach to an autonomous grasping/exploration strategy enabling robots to gain more autonomy in the domain of haptic learning.

ACKNOWLEDGMENT

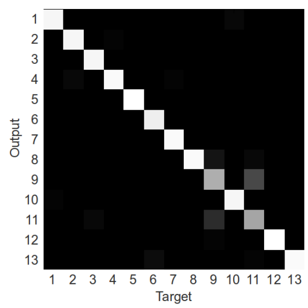
This work was supported by the German Research Foundation (DFG) within the Collaborative Research Center SFB 588 on “Hu-

(a) Set for Gripper

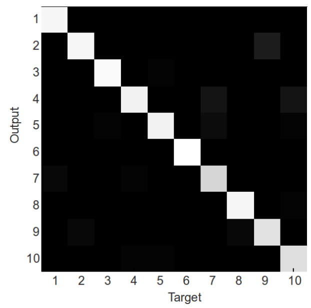
(b) Set for SDH-2 Hand

(c) Set for ARMAR-IIIb Hand

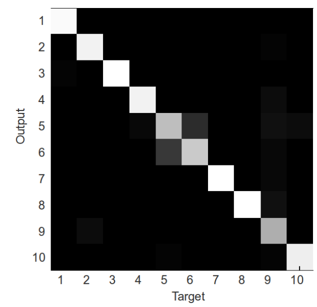
Figure 8: The object sets used for evaluation.



(a) Confusion matrix for Gripper set.



(b) Confusion matrix for SDH-2 set.



(c) Confusion matrix for ARMAR-IIIb Hand set.

Figure 9: Confusion matrices for the experiments.

manoid robots—learning and cooperating multimodal robots".

References

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

[2] P. Allen and K. Roberts. Haptic object recognition using a multi-fingered dextrous hand. In *Proc. of IEEE Int. Conf. on Robotics and Automation*, pages 342–247, 1989.

[3] T. Asfour, K. Regenstein, P. Azad, J. Schroder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann. Armar-III: An integrated humanoid platform for sensory-motor control. In *Proc. of the 6th IEEE-RAS International Conference on Humanoid Robots*, pages 169–175, Dec. 2006.

[4] I. Gaiser, S. Schulz, A. Kargov, H. Klosek, A. Bierbaum, C. Pylatiuk, R. Oberle, T. Werner, T. Asfour, G. Bretthauer, and R. Dillmann. A new anthropomorphic robotic hand. In *IEEE-RAS International Conference on Humanoid Robots*, 2008.

[5] G.Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[6] D. Göger, N. Gorges, and H. Wörn. Tactile Sensing for an Anthropomorphic Robotic Hand: Hardware and Signal Processing. In *Proc. of the IEEE Int. Conf. on Robotics and Automation, May 12 - 17, 2009, Kobe, Japan*, 2009.

[7] D. Göger and H. Wörn. A highly versatile and robust tactile sensing system. In *Proc. of the IEEE Conf. on Sensors*, Atlanta (GA), USA, 2007.

[8] Gorges, N. and Escaida Navarro, S. and Goeger, D. and Woern, H. Haptic Object Recognition Using Passive Joints and Haptic Key Features. In *2010 IEEE International Conference on Robotics and Automation*, 2010.

[9] M.-K. Hu. Visual Pattern Recognition by Moment Invariants. *IEEE Transactions on Information Theory*, 8(2):179–187, 1962.

[10] M. Johnsson and C. Balkenius. Experiments with proprioception in a self-organizing system for haptic perception. In *Towards Autonomous Robotic Systems*, pages 239–245, Aberystwyth, UK, 2007.

[11] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, January 1982.

[12] C. Magnanini and F. Zanichelli. Haptic object recognition with a dextrous hand based on volumetric shape representations. In *Proc. of the IEEE Int. Conf. on Multisensor Fusion and Integration, Las Vegas, NV*, pages 2–5, 1994.

[13] A. Schneider, J. Sturm, C. Stachniss, and W. B. Marc Reisert, Hans Burkhardt. Object identification with tactile sensors using bag-of-features. In *Proc. of the Int. Conf. on Intelligent Robot Systems, St. Louis, USA*, 2009.

[14] Schunk. SCHUNK GmbH & Co. KG Site: www.schunk.com. retrieved 2011/09/19.

[15] S. Takamuku, A. Fukuda, and K. Hosoda. Repetitive grasping with anthropomorphic skin-covered hand enables robust haptic recognition. In *Proc. of Int. Conf. on Intelligent Robot Systems, Nice, France*, pages 3212–3217, 2008.

[16] L. van der Maaten. Matlab toolbox for dimensionality reduction, November 2010. Accessed 06/16/2011.

[17] K. Weiss and H. Wörn. The working principle of resistive tactile sensor cells. In *Proc. of the IEEE Int. Conf. on Mechatronics and Automation, Canada*, 2005.

[18] Weiss Robotics. Tactile sensor module, type: DSA 9335. Accessed 10/09/2009.