2014 14th IEEE-RAS International Conference on
Humanoid Robots (Humanoids)
November 18-20, 2014. Madrid, Spain

Tactile Object Recognition using Deep Learning and Dropout

Alexander SCHMITZ, Yusuke BANSHO, *Kuniaki NODA,
*Hiroyasu IWATA, *Tetsuya OGATA and **Shigeki SUGANO
*Member, IEEE, **Fellow, IEEE

Abstract— **Recognizing grasped objects with tactile sensors is beneficial in many situations, as other sensor information like vision is not always reliable. In this paper, we aim for multimodal object recognition by power grasping of objects with an unknown orientation and position relation to the hand. Few robots have the necessary tactile sensors to reliably recognize objects: in this study the multifingered hand of TWENDY-ONE is used, which has distributed skin sensors covering most of the hand, 6 axis F/T sensors in each fingertip, and provides information about the joint angles. Moreover, the hand is compliant. When using tactile sensors, it is not clear what kinds of features are useful for object recognition. Recently, deep learning has shown promising results. Nevertheless, deep learning has rarely been used in robotics and to our best knowledge never for tactile sensing, probably because it is difficult to gather many samples with tactile sensors. Our results show a clear improvement when using a denoising autoencoder with dropout compared to traditional neural networks. Nevertheless, a higher number of layers did not prove to be beneficial.**

I. Introduction

In this paper we study object recognition with tactile sensors. According to the definition we use, tactile sensing includes distributed sensors in the skin as well as proprioceptive sensing, if the proprioceptive sensors are used to gain information about the outside world [1]. Multimodal tactile object recognition is useful in cases where vision fails and the proprioceptive information alone is insufficient for object recognition. In the past, various methods have been used for tactile object recognition, but many of them can be applied only to special cases. One possibility is to obtain a 3D point cloud of the object through repeatedly touching it, but a rather high amount of contacts at tightly controlled positions are required to explicitly model the object shape [2][3][4]. Other studies, for example [5][6][7][8][9][10], have focused on material based object recognition, but these approaches require exploratory actions different from power grasping behavior, for example knocking on the surface or sliding the sensors over the surface.

In this paper, we consider tactile object recognition through power grasps, as we try to avoid the necessity of special actions for tactile object recognition. Multimodal information (proprioceptive and distributed sensors in the skin) has been shown to be beneficial for robust tactile object recognition in this case, but only few multifingered robot hands with distributed sensors in the skin exist, and consequently few studies on multimodal tactile object recognition with multifingered hands have been published (a detailed review will be provided in the next section). Instead, in previous studies often grippers instead of multifingered hands were used. Even when mutifingered hands were used, flat sensor arrays were employed. Planar sensors are easier to produce, while curved sensors allow for more human like hands, which enable for example in-hand manipulation like in [11], but less information is obtained compared to using planar sensors [12], therefore adding difficulty to the recognition task. Another common limitation in previous studies is that they require the object to be oriented in a certain way relative to the hand. Instead, this paper uses the information from several grasps in unknown positions and orientations. The objects are allowed (and indeed expected) to move between grasping actions.

An intricate recognition task like this requires advanced machine learning algorithms in order to achieve high recognition rates. Previous studies on tactile object recognition have employed feature learning techniques like principal component analysis (PCA) and self organizing maps (SOM). The current paper explores the possibility of using deep learning and dropout for tactile object recognition and compares the recognition rates to previously employed machine learning techniques.

The contributions of this paper are therefore twofold:

1. Challenging tactile object recognition is attempted, which we consider to be closer than previous work to the task that is performed by humans in many situations. In particular, a multifingered hand that provides multimodal sensory information and power grasps are used to recognize the objects. The surface of the finger phalanges and the palm, and therefore the tactile arrays, are curved. Furthermore, the information from several grasps is used together, but the orientations and positions of the grasps are unknown. Moreover, a challenging set of objects has to be recognized, for example our object set includes 5 similarly shaped bottles.

2. In this paper, we compare methods that have been used in the past for tactile object recognition to novel

Alexander Schmitz is with the Sugano Lab, School of Creative Science and Engineering, Waseda University, Okubo 2-4-12, Shinjuku, Tokyo, 169-0072, Japan (e-mail: schmitz@aoni.waseda.jp)

Yusuke Bansho is with the Sugano Lab, School of Creative Science and Engineering, Waseda University, 41-204B(room), Kikuicho 17, Shinjuku, Tokyo, 162-0044, Japan (e-mail: y_bansho@sugano.mech.waseda.ac.jp)

Kuniaki Noda and Tetsuya Ogata are with the Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo 169-8555, Japan (e-mail: kuniaki.noda@akane.waseda.jp, ogata@waseda.jp)

Hiroyasu Iwata is with the Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University (e-mail: jubi@waseda.jp)

Shigeki Sugano is with the Department of Modern Mechanical Engineering, School of Creative Science and Engineering, Waseda University, Ookubo 3-4-1, Shinjuku, Tokyo, 169-8555, Japan, (phone: 81-3-5286-3264, Fax: 81-3-5272-0948, e-mail: sugano@waseda.jp)

methods. In particular, we compare the results when using shallow neural networks (with or without PCA) to the results when using deep and shallow neural networks with dropout. Deep learning has been rarely used in robotics (some of the few examples are [13][14][15]). Moreover, to our best knowledge this is the first paper that uses deep learning and dropout for tactile object recognition (and data coming from the tactile sensory domain in general). For our dataset, we could show a clear improvement with those novel machine learning techniques, in particular we increased the recognition rate from about 68% to about 88%. Interestingly, even with a shallow neural network, the recognition rate could be increased by about 10% with pretraining alone, or even by about 20% when only using dropout, respectively.

A. Overview of the Rest of this Paper

In Section II, first related work on object recognition with tactile sensors is given. Subsequently, a short introduction to deep learning and dropout is provided. In Section III, the TWENDY-ONE hand is presented. This section also describes how the data is collected, which sensors are used, and how the information from several grasps is combined. Section IV shows the results when using a shallow artificial neural network (ANN), with and without PCA; the recognition rate when using 1 or 4 grasps as input, and the results with and without the distributed tactile sensor are compared. Section V describes the results when using deep learning techniques, such as different number of hidden layers, and dropout. In Section VI conclusions are drawn and possible future work is presented.

II. Related Work

A. Previous Research on Object Recognition with Distributed Skin Sensors

Here we provide an overview of related work that used distributed skin sensors for object recognition, with a particular focus on the employed methods. In [16] features like point/line/area contact and movements of those contact classes are extracted by rolling objects of different shapes (cube, sphere, cone …) over a tactile array, which is bigger than the objects, and thereby simplifying the task. In [17][18][19] a tactile array is pressed from the top onto different objects. In [17][18], features are extracted with local principal component analysis (local PCA) from the time series of 2D pressure profiles, and a Local Linear Map (LLM, related to self-organizing maps) is used for classification. The same setup for data gathering is used in [19], but 51 different hand-designed features are extracted (value maximum taxel, mean taxel, position minimum taxel …) and decision trees are used for categorization. For [17][18][19], different translations and rotations have to be trained explicitly, and again the objects are smaller than the sensor.

In [20] a SOM was used to differentiate grasped objects, and it could be shown that the additional information from the tactile sensors in the fingertips could aid in distinguishing the objects. In [21] the robot learned to label objects with 34 adjectives, like absorbent, bumpy or compact. For the distributed sensors in the fingertips PCA was performed, and the first two principal components were found to capture the majority of the information, as the data was highly coupled due to the spatial proximity of the sensors. Significant information for the classification comes from other sensors and special exploratory actions. In [22] a robot with tactile sensors in its forearm brushes against different objects. The tactile image is converted to a binary image, three features are extracted (maximum force, contact area, and contact motion) over 40 time steps, which are combined to a feature vector of length 120, then a low dimensional representation is computed with PCA (20 principal components), before a classification with a k-nearest neighbor algorithm. The objects are either recognized or assigned to categories: rigid or soft, moveable or fixed.

In [23] a gripper with two tactile arrays grasps objects. A set of features (tactile images) is obtained through unsupervised k-means clustering of training data. The approach is multimodal, as also proprioceptive information (the grasping width) is used. When using proprioception plus the data from the surface sensors, more objects could be recognized than by just using one of those sensing modalities. Moreover, the objects are larger than the sensor arrays, therefore multiple grasps are necessary, and the next grasp is selected to maximize the expected information gain. Even visually similar objects like a soft and hard ball could be differentiated in this way. On the downside, a gripper is used, only the vertical grasping position is controlled, the tactile arrays are flat, and no dynamic information is used (no time series data). Furthermore, objects with different orientations are recognized as different objects. Also in [24] a number of 2D pressure profiles, that cover only parts of the whole object, in unknown position and orientation of the object, are used for recognizing the object. Either a set of complex 3-D objects in simulation or a set of raised letters also with real sensors is recognized. From the tactile images different descriptors are extracted (for example SIFT or image moments), then with PCA the principal components covering 90% of the variance are extracted from those descriptors, then clustering (either k-means or Gaussian mixture model) is applied, and then the per-class histograms are built.

In [25][26] it was shown that repetitive grasping converges objects into a stable grasping posture, which results in reduced variance of the sensor input. In [25] a SOM was used to differentiate between objects, and the authors in [26] used a recurrent neural network to use the time series data. While the exact initial posture of the objects is unknown, the objects are always roughly aligned, and cylindrical and spherical objects are used, which facilitates the convergence into a stable state.

A multifingered hand with flat tactile sensors on the phalanges was used in [12]. A Bayes classifier was used for object recognition. For the tactile skin sensors, either binary contacts, image moments (mean-x, mean-y, principal axis, eccentrity) or the results of a PCA and SOM were used; the best recognition rate was achieved with the image moments. The object recognition was performed independently for tactile skin and kinesthetic sensors; the average of both was better than either alone. When fusing the sensor results of several palpations together, a higher recognition rate could be achieved (steadily increasing from 1 to 6 grasps). The

inclusion of passive degrees of freedoms, to enable a good contact of the flat sensors with the objects, enhances the object recognition (but this could also make fine manipulation with the hands difficult). The recognition rate could be considerably improved by using a neural network (NN) classifier instead of a Bayesian classifier [27]. To get tactile features, PCA and subsequent SOM is used, for the kinesthetic features SOM is used, and always the features of 4 grasps are concatenated to form the input for the NN.

The works most closely related to ours are [12] and [27], as they also use a multifingered hand and power grasps for tactile object recognition. While they have used more traditional machine learning techniques (for example PCA and shallow neural networks), the current paper explores the possibility of using other machine learning techniques. The current paper uses different hardware (curved sensors, more sensor modalities) and a larger set of objects has to be recognized. As also stated in [27], in general it is difficult to compare and reproduce the results of other researches, as different and unique hardware are used. Nevertheless, the current paper compares the results of using PCA and a simple neural network classifier, two standard methods that have been used with success in the past for tactile object recognition, to deep learning machine learning techniques with dropout.

B. *Deep Learning and Dropout*

The multimodal tactile information for tactile object recognition we use is high dimensional, which can be problematic for classification and dimensionality compression mechanisms can be beneficial. However, it is very difficult to manually integrate all of the information by identifying and extracting the sensory features from each of the sensory modalities that are indispensable for robust object recognition. As presented in the last subsection, in previous works, conventional dimensionality compression mechanisms such as SOM or PCA and classifiers like shallow neural networks (with one hidden layer) have been utilized. On the other hand, recently in the machine learning community deep learning has attracted increasing attention, because it can extract higher level representation of sensory inputs. Hinton et al. also demonstrated that the compressed representation acquired from the deep learning outperforms the precision of acquired compressed features with PCA. Hence, in this work, we applied a deep learning framework for the self-organizing of sensory features and the multimodal sensory fusion.

In 2006, a breakthrough in deep learning was achieved [28]. One of the central ideas was greedy layerwise unsupervised pre-training. In this way, it is possible to use more hidden layers than it was before. Essentially, iteratively one layer of weights at a time is added to the neural network, using the outputs of one layer as input for the next level. After the layerwise unsupervised pre-training, the set of layers can be either used to initialize a supervised predictor, or the final output of the unsupervised pre-training can be used as input for a standard supervised machine learning predictor, such as a neural network classifier. It also possible to use the layerwise procedure in a purely supervised setting [29]. Hinton originally used restricted Boltzmann machines (RBMs) and stacked them into a Deep Belief Network [28]. Alternatively, also auto-encoders can be used [30], which are conceptually closer to neural networks. A variant are denoising auto-encoders (DAE), in which the input for each layer is artificially corrupted, which makes the outcome more insensitive to noise in the input, and which is particularly useful in cases where a limited number of input samples is available. As this is the case in the current study, we will use stacked DAEs. While the noise can be added in many different ways, we will use the variant in which a certain percentage of the input for the next layer is set to zero, which is called dropout. It has been shown that dropout is also beneficial for shallow neural networks [31]. Dropout can be seen as an extreme form of bagging and can be used for regularization. A more detailed review of deep learning techniques can be found in [32], and we used the practical advice in [33] for choosing the parameters in our deep learning attempts. For implementing the stacked DAE, we use the DeepLearnToolbox for Matlab [34].

III. Experimental Setup

A. *Robot Hand*

The hand of the human symbiotic robot TWENDY-ONE was used to grasp objects and get tactile data. It is one of only few robotic hands that have both distributed force sensors on many parts of hands, 6-axis F/T sensors in the fingertips and joint angle measurements [35]. In total, 241 distributed tactile skin sensors cover the hand, as depicted in Fig. 1. Moreover, the hand has 16 degrees of freedom. The DIP and PIP joints of the index, middle and little finger are coupled, and the hand is actuated by 13 small electric motors integrated in the joints. The DIP and MP1 joints also include springs; there are no springs for the thumb. The hand is also covered with a soft silicone layer. The compliance of the robot hand increases the number of tactile sensors that are in contact with the object, even with a simple grasping strategy. The importance of the hand design, in particular compliance, for tactile object recognition is discussed in [12][25][26][36]. The hand is about 20 cm long and the palm is 10 cm wide.

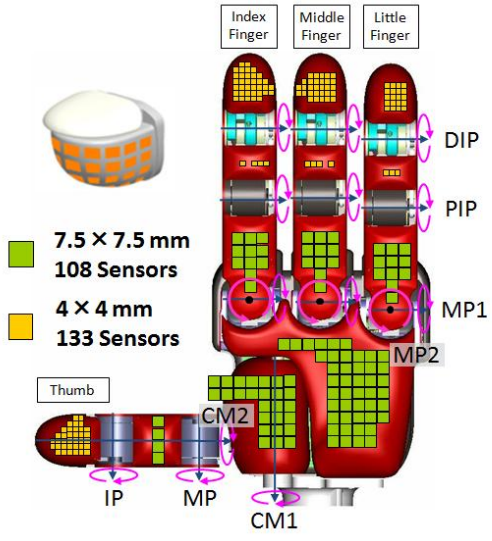

Fig. 1. The hand of the human symbiotic robot TWENDY-ONE.

Fig. 2. The objects used for the recognition task

Fig. 3. Example of 4 grasps

B. Data Collection

The robot grasped 20 different objects, which are depicted in Fig. 2. In order to test the power of our approach, some of the objects were chosen to be very similar, while others had a completely different shape or hardness. A simple strategy is used for grasping the objects and gathering sensor data. The object is handed to the robot in different orientations and translations, chosen randomly by the person handing the object to the robot. We aimed for handing each object to the robot in as many different orientations and translations as possible (four example grasps are depicted in Fig. 3). The palm of the robot either faced up or to the side. Each finger closed at the same time, until a threshold in the 6-axis F/T sensor or a maximum flexion was reached. The simple grasp strategy was sufficient, as the compliance due to the springs and the soft skin facilitates achieving a tight grip of the hand around the object, as discussed in the last subsection. In order to minimize the influence of the environmental factors on the classification, the data for each object was collected in several sessions, spread out over several days: for example, the robot performed several grasps for one object, then grasped the remaining objects, and then gathered again data for the first object, and so on. Furthermore, in order to counteract any drift in the sensors after the initial calibration, the sensor readings before each palpation were subtracted from the sensor data 1 second after all motors stopped moving, and this data is used as one grasping sample. For the current study, we ignored the other time series data. The robot grasps each object at least 20 times.

For each grasp, 312 values are recorded: 241 distributed tactile sensors, four 6-axis F/T sensors, 13 motor angles, 13 reference angles, 13 motor currents, and 8 spring displacements. We want to point out that the response of the distributed tactile sensors was very non-uniform from one sensor to the next at the time of the data gathering. We only removed 7 tactile sensor channels that had completely uniform sensor readings in all experiments (basically not measuring anything), with a resulting dimension of 305 for each grasp. All the remaining sensor data was left deliberately inside the data set, in order to test the stability of the learning algorithm to noisy input. Furthermore, it is not trivial to say which channels are pure noise, and which contain information. All inputs were normalized (0 mean, 1 standard deviation). From the normalized output, outliers below -2 and above 2 were set to -2 and 2, respectively.

C. Combining grasps

As presented in Section II, the information from several grasps combined can increase the recognition rate [12][23][27]. In this paper, the recognition rate when using 1 grasp to using the concatenated sensor data of 4 grasps as input for the classifier will be compared. When the data from 4 grasps is concatenated, the resulting input has a dimension of 1220. For training and testing, we use cross-validation: in our case, 4 grasps for each object are used for the test set, and the remaining grasps are used for the training set. For the test set, the four grasps for each object are used in all possible 24 possible permutations as input. Therefore, for 20 objects, the test set has 480 instances. For the training set, the grasps are either combined in different combinations, or permutations. To limit the number of instances in the training set, for combinations, a maximum of 2 grasps are shared across each pair of instances in the training set, similar to [27]. For permutations, the combinations are also used in different permutations. Therefore, for combinations, the training set has about 5000 instances and for permutations about 120000 instances.

IV. Object Recognition with a Shallow ANN and PCA

Initially, a simple neuronal network with one hidden layer was used, which was implemented with the patternnet of the neural network toolbox that is part of Matlab. First, the information from one grasp was used as the input. Different sets of sensors (for example, with or without the skin sensors, see Table I) were used as input for the classifier. We performed a random search to find optimized hyperparameters. As a result, we used a hidden layer size of 42; more neurons did not improve the results, but less neurons decreased the recognition rate. We used the default transfer function (tansig, mathematically equivalent to tanh) and default values for the other training parameters; other initial learning rates and other transfer functions did not improve the recognition rate. 70% of the samples were used for training, 15% as validation set, and 15% as testset. For each different input, the net was trained 30 times with a randomized split, and the average recognition rate is shown in Table I. From the results it can be seen that all sensors together are better than any of them alone.

Table II shows the results when using the concatenated input from 4 grasps as input. After performing a manual search, the hidden layer was chosen to have a size of 1000, and again the default values for the other training parameters were used. The data from 4 random grasps for each object was used for the testset (as we will show in Section V, those 4 grasps turned out by chance to be about averagely hard to

TABLE I. RESULTS WITH SIMPLE ANN AND 1 GRASP

DATASET	RECOGNITION RATE (%)
All data	49.3
Without skin sensors	41.0
Only 6 axis F/T	38.0
Motor angles + reference angles + currents + spring displacements	31.5
Only skin sensors	17.0

learn compared to other grasp sets) and the rest were used for training, with 15% of the training set used as a validation set (the results were nearly the same if no validation set was used). In general, the recognition rate could be drastically improved compared to when using only 1 grasp, especially when using the permutations. Again, if the input includes the distributed skin sensors, the results are better than without them. We concluded that it is beneficial to use the distributed tactile sensors, that 4 grasps are better than 1 grasp, and that permutations are better than combinations.

We also performed a PCA of all data with permutations and extracted the first 297 principal components which cover 90% of the variance. The result with the PCA was nearly identical to the case without it.

TABLE II. RESULTS WITH SIMPLE ANN AND 4 GRASPS

DATASET	RECOGNITION RATE (%)
All data / Permutations	67.6
All data / Combinations	55.0
Without tactile data / Permutations	61.7
Without tactile data / Combinations	49.3
Only tactile data / Permutations	39.7
Only tactile data / Combinations	19.2
PCA (297 principal components)	67.8

V. OBJECT RECOGNITION WITH DEEP LEARNING AND DROPOUT

We picked a random grasp set (the same 4 gasps for each object that we used in the last section) as the testset and optimized the learning hyperparameters. The best parameters can be seen in Table III. Here, after the layerwise unsupervised pre-training, the set of layers was used to initialize the supervised predictor. The maximum recognition rate we got was 85.41%. This was clearly an improvement over the results in the last section. The results with 297 principal components were nearly identical to the results with the original data, and we opted to use the original data. Also the following parameters did not affect the overall results (depending on the other factors, leading to a marginally higher or lower recognition rate): a lower initial learning rate for either to the supervised or unsupervised learning, an overcomplete representation in the first hidden layer (2000 neurons in the first hidden layer), or using more pre-training epochs. Using the hyperbolic tangent instead of sigmoid as the activation function, a higher initial learning rate, or adding weight penalty, decreased the recognition rate or made the learning slower. More than 3 hidden layers without pre-training made learning impossible (recognition rate 5%). Furthermore, it is worth mentioning that we saw no overfitting when we monitored the recognition rate of the testset in parallel to the recognition rate of the trainingset.

TABLE III. BEST PARAMETERS

PARAMETER	VALUE
Number of layers	5
Layer size (including input and output)	1220-800-600-400-200-100-20
Activation function	sigmoid
Pretraining epochs	4
Pretraing learning rate	1
Pretraining dropout fraction	50%
Supervised training epochs	20
Supervised learning rate	1
Supervised dropout fraction	50%

Interestingly, there was no clear trend that more layers are better. We validated this result by using 3 random combinations of grasps for each object as testset (the remaining grasps were used as training set, respectively). The hidden layer size for one hidden layer was 500, for two hidden layers 500-100, for three hidden layers 700-200-100, for four layers 800-600-400-200, for 5 layers 800-600-400-200-100, for 6 hidden layers 1000-800-600-400-200-100, and for 10 hidden layers 1000-900-800-700-600-500-400-300-200-100. The other parameters were as in Table III. Indeed there was no clear trend that more layers improve the recognition performance, as can be seen in Fig. 4.

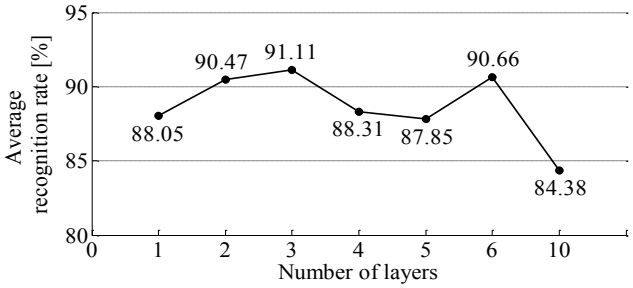

Fig. 4. Recognition rate for different number of hidden layers.

On the contrary, pre-training and dropout proved to be important for the recognition rate, as can be seen in Table IV. In this case, we used only one hidden layer (hidden layer size 500), and the other parameters were like before, if not stated otherwise. Either dropout or pre-training increases the recognition rate by more than 10%. The most single important change that can be made is to use dropout during the supervised training. A lower (20%) or higher (70%) dropout fraction was not beneficial. As we were just using 1 hidden layer, no pre-training and no dropout during the supervised training is the same case as in Section IV, and indeed the recognition rate is similar (69%) when using the

same grasps for the testset as in Section IV, which we consider to be a strong indiction that the two libraries we use (Neural Network Toolbox and DeepLearn Toolbox) are indeed comparable. Also with 3 hidden layers, without pre-training and no dropout during the supervised training, the recognition rate was nearly identical (68.5%). Using more hidden layers, with pre-training, but without dropout during the pre-training and the supervised training, achieves a recognition rate similar to the corresponding case for one hidden layer, which shows again that using more hidden layers is not beneficial in our case.

TABLE IV. RECOGNITION RATE IN DIFFERENT SCENARIOS

Pretraining dropout fraction in %	Supervised dropout fraction in %	RECOGNITION RATE (%)
No pretraining	0	64.7
No pretraining	50	86.0
0	0	75.8
50	0	81.2
0	50	86.0
50	50	88.3

We also tested the case in which the output of the unsupervised pre-training was used as input for the supervised predictor. In this case, the network structure was 1220-500-500-20, and the first hidden layer was trained purely unsupervised, and the top two purely supervised. In general, using the output of the pre-training as input for the supervised predictor has the benefit that the results of the pre-training can be used for other tasks as well, which makes this conceptually closer to feature learning. When using the same 3 random grasping sets as before as test sets, the recognition rate is 86.71%. This results are similar to the ones when using the unsupervised pre-training to initialize the supervised training (recognition rate: 88.3%), and indeed also similar to the ones of the purely supervised training (recognition rate: 86%).

As a final evaluation, we tested 76 random grasp combinations (not including the 4 grasps we used for hyperparameter optimization) with 1 hidden layer, otherwise using the same parameters as in Table III. Here, the result of the unsupervised pre-training was used to initialize the supervised predictor. The average recognition rate was 88.07%, with a standard deviation of 6.33. Fig. 5. shows the combined confusion matrix of those 76 grasp combinations. The most mixed up objects were the following: The coke bottle got recognized as the blue, trapezoidal bottle, or as the orange. The green cup got recognized as the Orangina bottle or the white cup. The container with hearts got recognized as the Orangina bottle. The white box got recognized as the lunch box, and the 4 sided Rubik's Cube got recognized as the lunch box. While some of those confusions are between objects that are similar, others objects are less similar (in particular: coke bottle – orange; container with hearts – Orangina bottle; green cup – Orangina bottle).

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

In this paper, we presented tactile object recognition with a multifingered hand. As expected, the recognition rates were higher when using the multimodal sensor information from all sensors compared to when using only selected sensor modalities; the combined information from 4 grasps was higher than using only 1 grasp; permutations are better than combinations. On the other hand, using PCA did not increase the recognition rate. Interestingly, while more hidden layers only slightly improved the results, the recognition rate could be drastically increased by pre-training and/or in particular by using dropout while training. We thereby achieved a recognition rate of about 88%.

B. Future Work

We consider tactile object recognition to be an open problem, with many approaches yet to be compared. Better hardware as well as more advanced software will enable higher recognition rates. For example, we plan to use time series data for future work, which probably provides additional information, and also regrasping like in [26].

ACKNOWLEDGMENT

Part of this research was supported by JSPS Grant-in-Aid for Scientific Research (S) No. 25220005. Part of this research was supported by the Research Institute for Science and Engineering, Waseda University. Part of this research was supported by the Postdoctoral Fellowship of the Japan Society for the Promotion of Science (JSPS). Part of this research was supported by JST PRESTO "Information Environment and Humans" and MEXT Grant-in-Aid for Scientific Research on Innovative Areas "Constructive Developmental Science" (24119003).

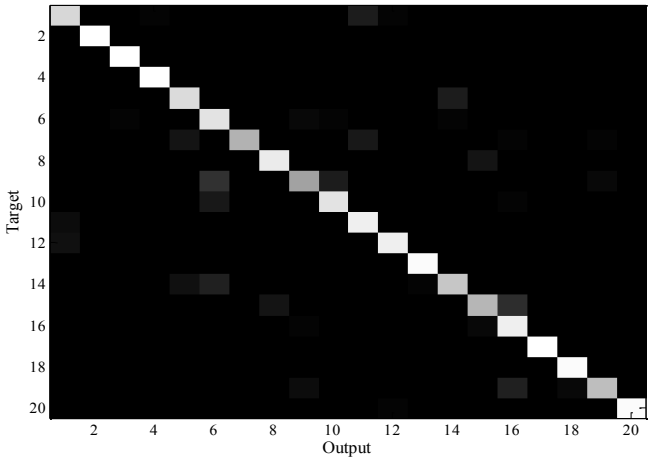

Fig. 5. Confusion matrix of 76 random grasp combinations that were used as testset, with denoising and pretraining. The twenty objects are (see also Fig. 2.): 1. ball white, 2. lunch box black, 3. bottle green tea hexagonal, 4. bottle transparent square, 5. bottle blue trapezoidal, 6. bottle Orangina, 7. bottle coke, 8. cookie box yellow hexagonal, 9. cup green, 10. cup white, 11. fruit orange, 12. fruit grape, 13. can, 14. container with hearts, 15. box white, 16. box red, 17. teddy bear, 18. cookie box orange, 19. four sided Rubik's Cube, 20. funnel.

REFERENCES

[1] R. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing: From humans to humanoids," IEEE Transactions on Robotics, vol. 26, no. 1, 2010

[2] A. Bierbaum, M. Rambow, T. Asfour and R. Dillmann, "Grasp Affordances from Multi-Fingered Tactile Exploration using Dynamic Potential Fields", Proc. IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2009

[3] P.K. Allen, "Integrating Vision & Touch for Object Recognition Tasks", International J. Robotics Research, Vol. 7, pp. 15–33, 1988

[4] M. Meier, M. Schopfer, R. Haschke and H. Ritter, "A Probabilistic Approach to Tactile Shape Reconstruction," IEEE Transactions on Robotics, vol.27, no.3, pp.630-635, 2011.

[5] J.M. Romano, T. Yoshioka and K.J. Kuchenbecker, "Automatic Filter Design for Synthesis of Haptic Textures from Recorded Acceleration Data", Proc. IEEE International Conference on Robotics and Automation (ICRA), 2010

[6] N. Jamali and C. Sammut, "Material Classification by Tactile Sensing Using Surface Textures", Proc. ICRA, 2010

[7] J. Windau and W. Shen, "An Inertia-Based Surface Identification System", Proc. ICRA, 2010

[8] S. Chitta, M. Piccoli and J. Sturm, "Tactile object class and internal state recognition for mobile manipulation", Proc. ICRA, 2010.

[9] T. Sugaiwa, G. Fujii, H. Iwata and S. Sugano, "A Methodology for Setting Grasping Force for Picking up an Object with Unknown Weight, Friction, and Stiffness," Proc. Humanoids, 2010

[10] Kroemer, O.; Lampert, C.H.; Peters, J. (2011). Learning Dynamic Tactile Sensing with Robust Vision-based Training, IEEE Transactions on Robotics (T-Ro), 27, 3, pp.545-557

[11] K. Kojima, T. Sato, A. Schmitz, H. Arie, H. Iwata and S. Sugano, "Sensor Prediction and Grasp Stability Evaluation for In-Hand Manipulation," Proc. IROS 2013

[12] N. Gorges, S.E. Navarro, D. Göger and H. Worn, "Haptic object recognition using passive joints and haptic key features," Proc. ICRA, 2010

[13] I. Lenz, H. Lee and A. Saxena, "Deep Learning for Detecting Robotic Grasps," Proc. Robotics: Science and Systems (RSS), 2013

[14] K. Noda, H. Arie, Y. Suga and T. Ogata, "Multimodal integration learning of object manipulation behaviors using deep neural networks," Proc. IROS, 2013.

[15] K. Noda, H. Arie, Y. Suga and T. Ogata, "Multimodal integration learning of robot behavior using deep neural networks," Robotics and Autonomous Systems, Volume 62, Issue 6, 2014

[16] R.A. Russell, "Object recognition by a 'smart' tactile sensor", Proc. Australian Conference on Robotics and Automation, 2000.

[17] G. Heidemann and M. Schopfer, "Dynamic tactile sensing for object identification," Proc. ICRA, 2004.

[18] M. Schopfer, H. Ritter and G. Heidemann, "Acquisition and Application of a Tactile Database," Proc. ICRA, 2007.

[19] M. Schopfer, M. Pardowitz and H. Ritter, "Using entropy for dimension reduction of tactile data," Proc. of International Conference on Advanced Robotics(ICAR), 2009.

[20] L. Jamone, G. Metta, F. Nori and G. Sandini, "James: A Humanoid Robot Acting over an Unstructured World," Proc. Humanoids, 2006

[21] V. Chu, I. McMahon, L. Riano, C.G. McDonald, Q. He; J. Martinez Perez-Tejada, M. Arrigo, N. Fitter, J.C. Nappo, T. Darrell and K.J. Kuchenbecker, "Using robotic exploratory procedures to learn the meaning of haptic adjectives," Proc. ICRA, 2013

[22] T. Bhattacharjee, J.M. Rehg and C.C. Kemp, "Haptic Classification and Recognition of Objects Using a Tactile Sensing Forearm," Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012

[23] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt and W. Burgard, "Object identification with tactile sensors using bag-of-features," Proc. IROS, 2009

[24] Z. Pezzementi, E. Plaku, C. Reyda and G.D. Hager, "Tactile-Object Recognition From Appearance Information," IEEE Transactions on Robotics, vol.27, no.3, pp.473-487, 2011

[25] S. Takamuku, A. Fukuda and K. Hosoda, "Repetitive grasping with anthropomorphic skin-covered hand enables robust haptic recognition," Proc. IROS, 2008

[26] K. Hosoda and T. Iwase, "Robust haptic recognition by anthropomorphic bionic hand through dynamic interaction," Proc. IROS, 2010

[27] S.E. Navarro, N. Gorges, H. Worn, J. Schill, T. Asfour. and R. Dillmann, "Haptic object recognition for multi-fingered robot hands," Proc. of IEEE Haptics Symposium (HAPTICS), 2012

[28] G.E. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks." Science 313.5786 (2006): 504-507.

[29] Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, "Greedy layer-wise training of deep networks," Proc. Neural Information Processing Systems (NIPS), 2007.

[30] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion." in The Journal of Machine Learning Research, 11: 3371-3408, 2010

[31] P. Baldi and P. Sadowski, "The dropout learning algorithm", Artificial Intelligence, Volume 210, Pages 78-122, May 2014

[32] Y. Bengio, A. Courville and P. Vincent, "Representation Learning: A Review and New Perspectives", Arxiv, 2012.

[33] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," ArXiv, 2012

[34] R.B. Palm, "Prediction as a candidate for learning deep hierarchical models of data", 2012

[35] H. Iwata and S. Sugano, "Design of Human Symbiotic Robot TWENDY-ONE," Proc. ICRA, 2009

[36] L. Natale, G. Metta and G. Sandini, "Learning haptic representation of objects," Proc. of International Conference of Intelligent Manipulation and Grasping, 2004