Article

Using 3D Convolutional Neural Networks for Tactile Object Recognition with Robotic Palpation

Francisco Pastor*, Juan M. Gandarias, Alfonso J. García-Cerezo and Jesús M. Gómez-de-Gabriel

Robotics and Mechatronics Group, Telerobotic and Interactive Systems Laboratory, University of Málaga, Spain.
fpastor@uma.es (F.P.), jmgandarias@uma.es (J.M.G.), ajgarcia@uma.es (A.G.C.), jesus.gomez@uma.es (J.G.G.)
* Corresponding author

Version January 21, 2024 submitted to Sensors

Abstract: In this paper, a novel method of active tactile perception based on 3D neural networks and a high-resolution tactile sensor installed on a robot gripper is presented. A haptic exploratory procedure based on robotic palpation is performed to get pressure images at different grasping forces that provide information not only about the external shape of the object but also about its internal features. The gripper consists of two underactuated fingers with a tactile sensor array in the thumb. A new representation of tactile information as 3D tactile tensors is described. During a squeeze-and-release process, the pressure images read from the tactile sensor are concatenated forming a tensor that contains information about the variation of pressure matrices along with the grasping forces. These tensors are used to feed a 3D Convolutional Neural Network (3D CNN) called 3D TactNet, which is able to classify the grasped object through active interaction. Results show that 3D CNN performs better, and provide better recognition rates with a lower number of training data.

Keywords: Tactile perception; Robotic palpation; Underactuated grippers; Deep learning

1. Introduction

Recent advances in Artificial Intelligence (AI) have brought the possibility of improving robotic perception capabilities. Although most of them are focused on visual perception [1], existing solutions can also be applied to tactile data [2–4]. Tactile sensors measure contact pressure from other physical magnitudes, depending on the nature of the transducer. Different types of tactile sensors [5–9] have been used in robotic manipulation [10,11] for multiple applications such as slippage detection [12,13], tactile object recognition [14,15] or surface classification [16,17] among others.

Robotic tactile perception consists in the integration of mechanisms that allow a robot to sense tactile properties from physical contact with the environment along with intelligent capacities to extract high-level information from the contact. The sense of touch is essential for robots the same way as for human beings to performing both simple and complex tasks such as object recognition or dexterous manipulation [18–20]. Recent studies focused on the development of robotic systems that behaves similar to humans, including the implementation of tactile perception capabilities [21,22]. However, tactile perception is still a fundamental problem in robotics that has not been solved so far [23]. Also, there are multiple applications, not limited to classic robotic manipulation problems, that can benefit from tactile perception such as medicine [24], food industry [3], or search-and-rescue [4] among others.

Many works related to tactile perception use pressure images after the interaction [25], which means that the interaction is considered static or passive. However, tactile perception in the real world is intrinsically active [26]. A natural or bio-inspired haptic Exploratory Procedure (EP) for perceiving pressure or stiffness of an object must consider dynamic information [27]. According to [28], the haptic attributes that can be perceived depends on the EP.

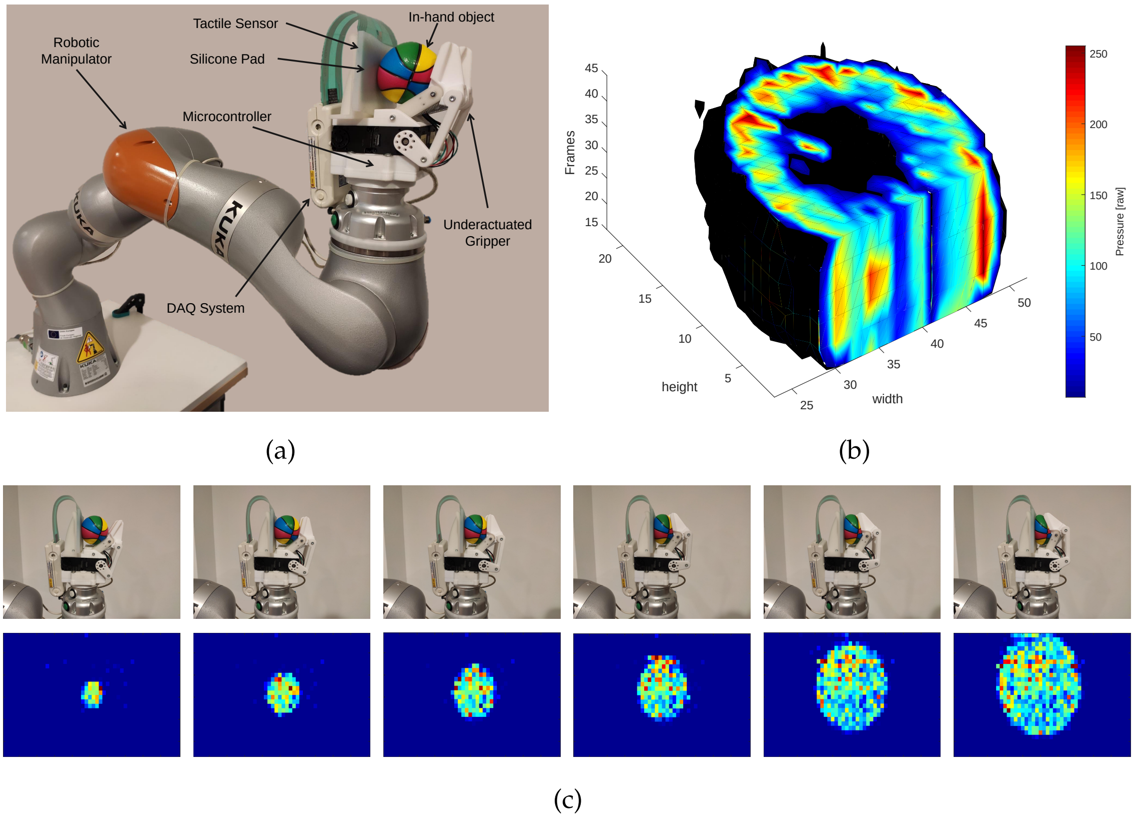


Figure 1. The full experimental system formed by a robotic manipulator, an underactuated gripper with a tactile sensor, and the control electronics (a), a 3D tensor representation of active tactile information when the gripper is grasping a squeezable ball (b), and a subset of pictures and their respective tactile images of a grasping sequence of another squeezable ball (c). In (b), the tensor is sectioned to show the intrinsic attributes and pressure variations of the grasped object.

survey on the concept of active tactile perception considering biological and psychological terms is presented in [29]. In this chapter, and according to [30], two approaches for tactile perception in robots are possible: perception for action, which means that the perceived information is used to guide the robot (i.e., dexterous manipulation and grasp control), and action for perception, which means that the robot explores the environment to collect data (i.e., active perception and haptic exploration). Hence, an active tactile perception approach can be defined as one in which the data are collected during an active EP using an active sensing approach (e.g., tactile sensing). This means that action and perception are not separated, and the robot collects dynamic data depending on the action, while this action is occurring. Therefore, although both static and dynamic tactile data are useful for many robotic applications, it can be considered that active perception is more faithful to the real sense of touch, and the information acquired using active tactile sensing reflects better the attributes of the grasped objects. Static pressure images only contain information about stiffness and shape of the object when a certain force is applied [14], while the changes of the pressure distribution over force, contain information about the variation of shape and stiffness during the whole EP [31]. This dynamic information allows us to distinguish both rigid and deformable objects [32].

This paper addresses the shortcomings mentioned above and is focused on the active tactile perception problem in robotics. A robotic palpation process with active tactile sensing, based on a squeeze-and-release motion for distinguishing grasped objects, both rigid and deformable, is presented (See Fig. 1). The robotic EP conceives a novel representation of dynamic tactile information based on sequences of pressure images and an AI method based on 3D Convolutional Neural Networks (3D CNNs) for active tactile perception. A tactile sensor array is integrated into the thumb of a gripper with two underactuated fingers to get sequences of tactile images. These sequences are represented as 3D tensors similar to Magnetic Resonance Imaging (MRI). However, in this case, 3D tactile tensors represent the variation of pressure distribution over applied force, whereas MRI contains information about cross-sectional images of internal structures and organs over distance. Although the type of information contained in MRIs and 3D tactile tensors is different, methods such as 3D CNNs used to process MRI information [33,34], might be used for tactile data with good results in this application as we explored in our previous work [35]. In this work, our preliminary study is expanded: A high-resolution tactile sensor has been integrated into a new gripper where the palpation process (e.g., the EP) is fully autonomous, so the robot controls the grasping force. As a result, not only objects with different elasticity are compared and classified, but also objects that contain internal inclusions and bags of objects which provide different pressure images each time, have been tested. In particular, 24 objects have been used: rigid, deformable, and in-bag; and the results are compared against 2D CNN-based methods. Altogether, the main contribution of this paper relates to the entire process of active tactile perception, considering the use of an underactuated, sensorized gripper to carry out the EP, and a 3D CNN for tactile perception.

The relevance of this contribution relies on different factors. First, the presented method achieves better performance in the discrimination problem for all kinds of objects, and in case that the number of classes increases, a lower number of training data is needed to obtain higher accuracy rates than classic 2D networks. Second, it is also shown that in case of misclassification, the resulting object class has almost indistinguishable physical features (e.g., soda cans of different capacities), where 2D CNNs, in the event of failure, give disparate output classes unrelated to the class of the grasped object.

This paper is structured as follows: In section 2, the current state-of-the-art related to this topic is introduced. In section 3, the underactuated gripper and the 3D CNN-based method used for tactile perception are described. The experimental protocol and results are explained in section 4, and a thorough and detailed discussion of our results in comparison with related works is presented in section 5. Finally, the conclusion and future research lines are exposed in section 6.

2. Related Work

Related works within the scope of tactile perception in robotics focus on tactile object-recognition from pressure-images, deep-learning methods based on CNNs, and active tactile perception.

2.1. Tactile object recognition

Two main approaches for tactile object recognition may be considered depending on the nature of the EP: On one hand, perceiving attributes from the material composition, which are typically related to superficial properties like roughness, texture, or thermal conductivity [36–38]. On the other hand, other properties related to stiffness and shape may also be considered for object discrimination [39–41]. Most of these works are based on the use of novel machine learning-based techniques. That way, different approaches can be followed, such as Gaussian Processes [42], k-Nearest Neighbour (kNN) [25], Bayesian approaches [43], k-mean and Support Vector Machines (SVM) [44] or Convolutional Neural Networks (CNNs) [45] among others. Multi-modal techniques have also been considered in [46], where they demonstrated that considering both haptic and visual information generally gives better results.

2.2. Tactile perception based on pressure images

Concerning the latter approach, most of the existing solutions in literature acquire data from tactile sensors, in the form of matrices of pressure values, analog to common video images [47]. In this respect, multiple strategies and methodologies can be followed. In [25] a method, based on *Scale Invariant Feature Transform* (SIFT) descriptors, is used as a feature extractor, and the kNN algorithm is used to classify objects by their shape. In [15], Luo et al. proposed a novel multi-modal algorithm that mixes kinesthetic and tactile data to classify objects from a 4D point cloud where each point is represented by the 3D position of the point and the pressure acquired by a tactile sensor.

2.3. CNNs-based tactile perception

One recent approach for tactile object discrimination consists of the incorporation of modern deep learning-based techniques [48,49]. In this respect, the advantages of Convolutional Neural Networks (CNNs) such as translational and rotational invariant property, enable the recognition in any pose [50]. A CNN-based method to recognize human hands in contact with an artificial skin has been presented in [51]. The proposed method benefits from the CNNs translation-invariant properties and is able to identify whether the contact is made with the right or the left hand. Apart from that, the integration of the dropout technique in deep learning-based tactile perception has been considered in [49], where the benefits of fusing kinesthetic and tactile information for object classification are also described, as well as the differences of using planar and curved tactile sensors.

2.4. Active tactile perception

In spite of the good results obtained by existing solutions in tactile object recognition, one of the main weaknesses is that most of these solutions only consider static or passive tactile data [25]. As explained, static tactile perception is not a natural EP to perceive attributes like pressure or stiffness [27]. Pressure images only have information about the shape and pressure distribution when a certain force is applied [14]. On the other hand, sequences of tactile images also contain information about the variation of shape (in the case of deformable objects [32]), stiffness and pressure distribution over time [31].

Time-series or sequential data are important to identify some properties. This approach has been followed in some works for material discrimination [52,53]. In [54], an EP is carried out by a robotic manipulator to get dynamic data using a 2D force sensor. The control strategy of the actuator is critical to apply a constant pressure level and perceive trustworthy data. For this purpose, a multi-channel neural network was used achieving high accuracy levels.

Pressure images obtained from tactile sensors have also been used to form sequences of images. In [3], a flexible sensor was used to classify food textures. A CNN was trained with sequences of tactile images obtained during a food-biting experiment in which a sensorized press is used to crush food, simulating the behavior of a mouth biting. The authors found that the results when using the whole biting sequence or only the first and last tactile images, were very similar because the food was crushed when a certain level of pressure was applied. Therefore the images before and after the break point were significantly different. For other applications, as it was demonstrated in [55], Three-Dimensional Convolutional Neural Networks (3D CNNs) present better performance when dealing with sequences of images than common 2D CNNs.

3. Materials and Methods

The experimental setup is composed by a gripper with a tactile sensor. The gripper, the representation of 3D tactile information, and the 3D CNN are described next.

3.1. Underactuated gripper

The active perception method has been implemented using a gripper with two parallel underactuated fingers and a fixed tactile-sensing surface (See Fig. 2). The reason for using an underactuated gripper is that this kind of gripper allows us to apply even spread pressure to the grasped objects, and the fingers could adapt to their shape, which is especially useful when grasping deformable or in-bag objects. In our gripper, each underactuated finger has two phalanxes with two DOF's θ_1 and θ_2, and a single actuator θ_a capable of providing different torque values τ_a. The values of the parameters of the kinematics are included in Table 1. A spring provides stiffness to the finger to recover the initial position when no contact is detected.

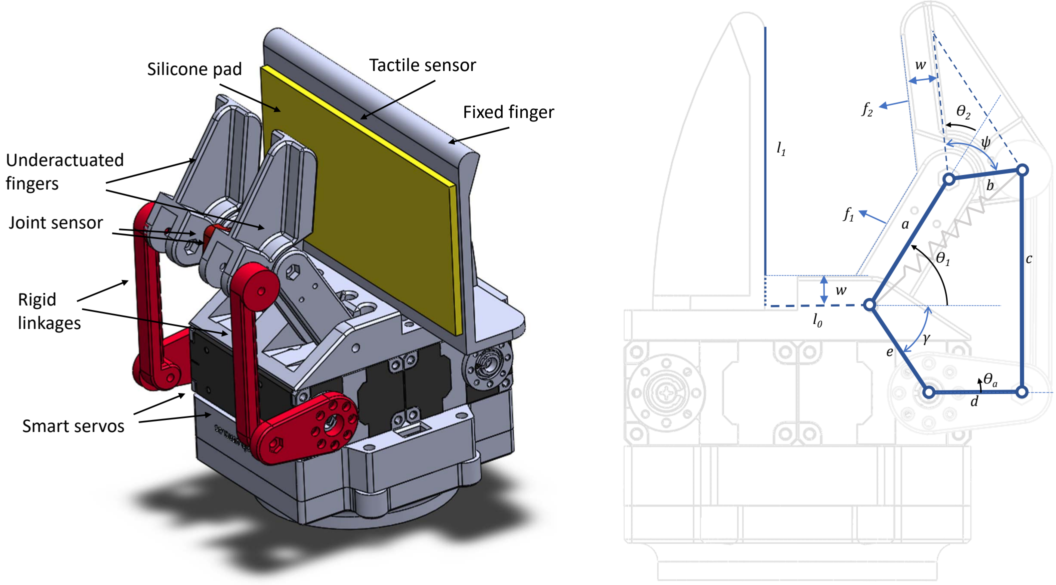


Figure 2. Gripper design (left) with two independent underactuated fingers and one fixed thumb with a tactile sensor covered with a silicone pad. The kinematic structure of the underactuated fingers (right) shows the five-bar structure with associated parameters and DOF's.

Two smart servos (Dynamixel XM-430 from ROBOTIS) have been used to provide different torques trough rigid-links, using a five-bar mechanical structure to place the servos away from the first joint. Thus, the relationship between τ_a and the joint torques (τ_1, τ_2) can be expressed as a transfer matrix T, and the computation of the Cartesian grasping forces (f_1, f_2) from the joint torques is defined by the Jacobian matrix $\mathbf{F} = \mathbf{J}(\boldsymbol{\theta})\boldsymbol{\tau}$.

The computation of those matrices requires knowledge of the actual values of the underactuated joints. For this reason, a joint sensor has been added to the second joint of each finger. The remaining joint can be computed as the actual value of the servo joint, which is obtained from the smart servos. Two miniature potentiometers (*muRata* SV01) have been used to create a special gripper with both passive adaptation and proprioceptive feedback.

The dynamic effects can be neglected when considering slow motions and lightweight fingers. This way, a kinetostatic model of the forces can be derived(1) as described in [56].

$$\mathbf{F} = \mathbf{J}(\boldsymbol{\theta})^{-T}\,\mathbf{T}(\boldsymbol{\theta})^{-T}\,\boldsymbol{\tau} \tag{1}$$

Although the actual Cartesian forces could be computed, each object with a different shape should require feedback control to apply the desired grasping forces. In order to simplify the experimental setup, an open-loop force control has been used for the grasping operations, where the actuation (PWM) of the DC motors of the smart servos follows a slow triangular trajectory from a minimum value (5%) to a maximum (90%) of the maximum torque of 1.4 N.m of each actuator. The resulting position of each finger depends on the actual PWM and the shape and impedance of each contact area.

Table 1. Parameter values for the kinematic model of the gripper with underactuated fingers.

Parameter	Value	Parameter	Value
a	40 mm	e	27.8 mm
b	20 mm	ψ	90°
c	60 mm	γ	56°
d	25 mm	w	10 mm
l_0	25 - 45 mm	l_1	70 mm

Table 2. Main features of the *Tekscan* 6077 tactile sensor.

Parameter	Value
Max. pressure	34 KPa
Number of tactels	1700
Tactels density	27.6 tactels/cm^2
Temperature range	$-40\,^{\circ}$C to $+60\,^{\circ}$C
Matrix height	53.3 mm
Matrix width	95.3 mm
Thickness	0.102 mm

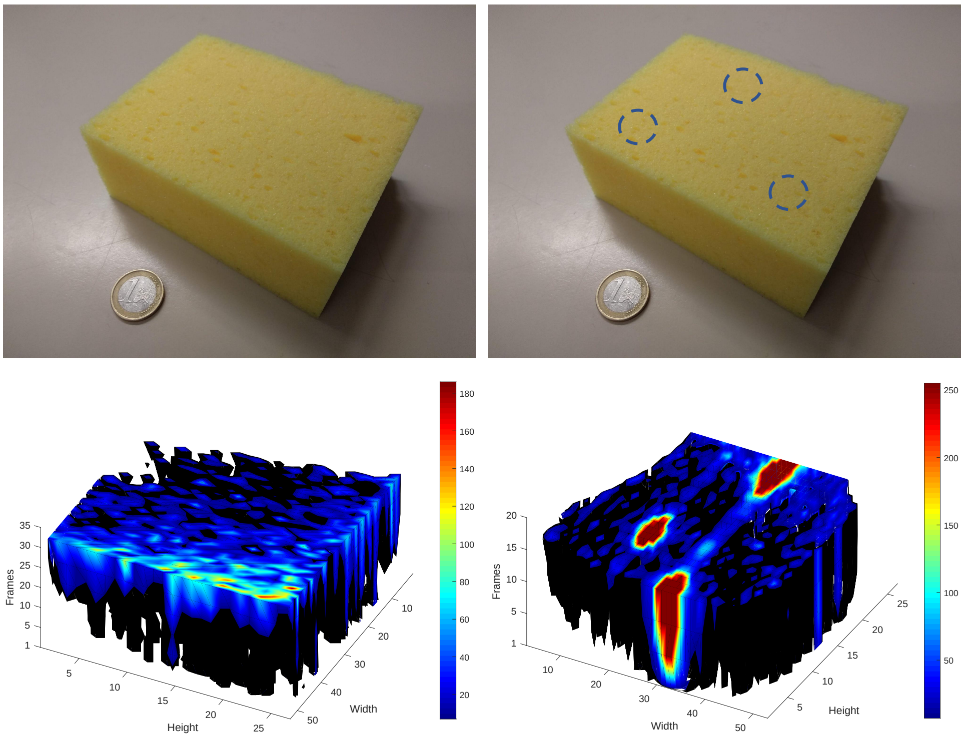


Figure 3. 3D tactile tensors (bottom) of the same sponge with and without hard inclusions (top). The inclusions become visible as grasping force increases, but cannot be seen in the picture of the sponge.

Finally, a microcontroller (Arduino Mega2560) has been used to acquire angles form the analog potentiometers and communicating with the smart servos in real-time, with a 50 ms period.

3.2. Tactile Sensor

A Teskcan (South Boston, MA, USA) sensor model 6077 has been used. This high-resolution tactile-array has 1400 tactels (also called taxels o sensels), with 1.3×1.3 mm size each. The sensor presents a density of 27.6 tactels/cm^2 distributed in a 28×50 matrix. The main features of the sensor are presented in Table 2. The setup includes the data acquisition system (DAQ) (see Fig. 1(a)), and the Tekscan real-time SDK.

A silicone pad of 3 mm has been added to the tactile sensor to enhance the grip and the image quality, especially when grabbing rigid objects. In particular, the Ecoflex$^{TM}00-30$ rubber silicone has been chosen due to its mechanical properties.

3.3. Representation of active tactile information

As introduced in section 1, a natural palpation EP to get information about the stiffness of an in-hand object is dynamic. In this respect, it seems evident that a robotic EP should also be dynamic so

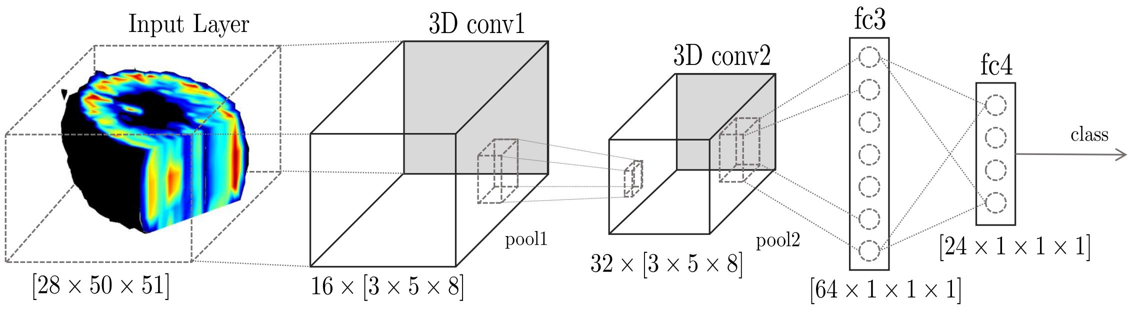


Figure 4. Architecture of TactNet3D, which is formed by 4 layers, the first two are 3D convolutional layers with kernel sizes $16 \times [5 \times 3 \times 8]$ and $32 \times [5 \times 3 \times 8]$ respectively, and two fully connected layers with 64 and 24 neurons respectively.

that the information acquired during the whole squeeze-and-release process describes the external and internal tactile attributes of an object.

The pressure information can be represented in multiple ways, commonly as sequences of tactile images. However, in this case, a more appropriate structure is in the form of 3D tactile tensors. An example of this type of representation is presented in 1(b), which is similar to MRI, except that in this case the cross-sectional images contain information about the pressure distribution at the contact surface, for different grasping forces.

To show the advantages of 3D tactile tensors, sectioned tensors of the same sponge, with and without hard inclusions, are shown in Fig. 3. The inclusions become perfectly visible as the grasping force increases.

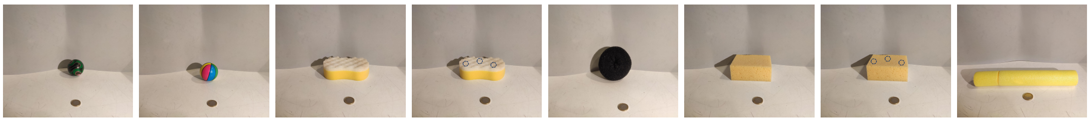
3.4. 3D TactNet

When using 3D tactile information, it is necessary to control the applied forces to obtain a representative pressure-images from a certain object. For 3D CNNs, each tensor has information about the whole palpation process. On the other hand, when dealing with soft or shape-changing objects, this operation is more challenging using 2D CNNs, as it would be necessary a high amount of training data, or selected data captured at optimal pressure levels, which also depends on the stiffness of each object.

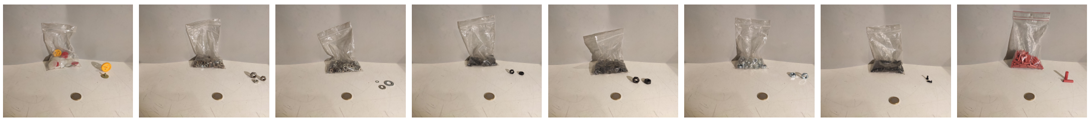
In previous works, we trained and validated multiple 3D CNNs with different structures and hyperparameters to discriminate deformable objects in a fully-supervised collection and classification process [35]. Here, although the classification is still supervised, the grasping and data collection processes have been carried out autonomously by the robotic manipulator. According to the results of our previous work, the 3D CNN with highest recognition rate, and compatible with the size of the 3D tensors read from our tactile sensor, was a neural network with four layers, where the first two were 3D convolutional, and the last two were fully connected layers. The network's parameters have been slightly modified to fit a higher number of classes and to adjust the new 3D tensor, which has a dimension of $[28 \times 50 \times 51]$.

The architecture of this network, called TactNet3D, is presented in Fig. 4. This network has two 3D convolutional layers ($\mathscr{C} = [\text{3D conv}_1\ , \text{3D conv}_2]$) with kernels $16 \times [3 \times 5 \times 8]$ and $32 \times [3 \times 5 \times 8]$ respectively, and two fully connected layers ($\mathscr{F} = [\text{fc}_3, \text{fc}_4]$) with 64 and 24 neurons respectively. Each convolutional layer also includes Rectified Linear Unit (ReLU), batch normalization with $\epsilon = 10^{-5}$, and max-pooling with filters and stride equal to 1. Besides, fc_3 incorporated a dropout factor of 0.7 to prevent overfitting. Finally, a softmax layer is used to extract the probability distribution of belonging to each class. The implementation, training and testing of this network has been done using the Deep Learning Toolbox in Matlab.

(a)



(b)



(c)

Figure 5. Pictures of the 24 objects used in experiments. Rigid objects (a), from left to right: bottle of coke, energy drink can, mouse 1, mouse 2, bottle of ice tea, skate wheel, soda can, and bottle of water. Deformable objects (b), from left to right: ball 1, ball 2, sponge rough, sponge rough with inclusions, sponge scrunchy, sponge soft, sponge soft with inclusions, and sponge pipe. In-bag objects (c), from left to right: gears, mixed nuts, mixed washer, M6 nuts, M8 nuts, M10 nuts, rivets and rubber pipes

4. Experimental Protocol and Results

This section presents the procedure for the dataset collection and the experiments. The dataset is conformed by 3 subsets of data: Rigid, deformable, and in-bag objects, which are described in more detailed below. Similarly, 4 experiments have been carried out to show the performance of the method and compare the results of dynamic and static methods: Experiment 1 for rigid objects, experiment 2 for deformable objects, experiment 3 for in-bag objects, and experiment 4 for the whole dataset.

4.1. Dataset

4.1.1. Collection process

The dataset collection process consists of capturing sequences of tactile images and creating a 3D tactile tensor. For this purpose, the underactuated gripper holds an object and applies incremental forces while recording images over the whole palpation process. Each object, depending on its internal physical attributes, has a unique tactile frame for each amount of applied force. The dataset collection has been carried out by the gripper, recording 51 tactile frames per squeeze. This process is made by the two active fingers of the gripper, which are moved by the two smart servos in torque control mode with incremental torque references. Finally, 1440 3D tactile tensors have been obtained, for a total of 24 objects with 60 tactile tensors each. In Fig. 1 (c), a grasping sequence is shown. The sequence at the top, from the left to the right, shows the grasping sequence due to the progressive forces applied by the underactuated gripper to the ball 2, and the sequence at the bottom, from the left to the right, shows the tactile images captured by the pressure sensor.

For machine learning methods, it is important to have the greatest possible variety in the dataset. In order to achieve this goal, the incremental torque is increased in random steps, so that the applied forces between two consecutive frames is different in each case. This randomness is also applied due to the intention to take a dataset that imitates the palpation procedure that could be carried out by a human, in which the exact forces are not known. Another fact that has been considered for the dataset collection process is that the force is applied to the object through the fingers of the gripper, therefore

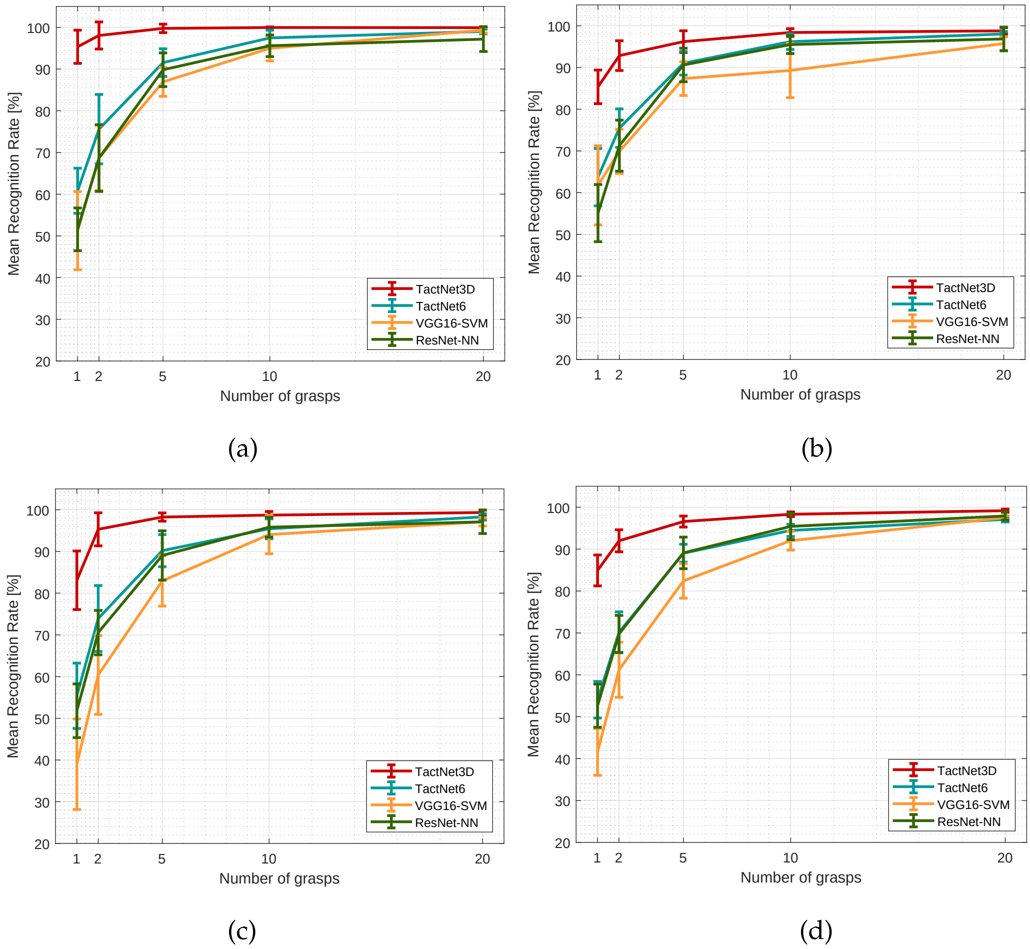


Figure 6. Experimental results of the experiment with rigid objects (a), deformable objects (b), in-bag objects (c), and all objects (d). Error bars represent the standard deviation σ of each recognition rate distribution over a 20 sample testing process.

non homogeneous pressure is exerted on the whole surface of the object. Therefore, in order to obtain all of the internal features of the objects, multiple grasps with random positions and orientations of the objects have been obtained.

4.1.2. Rigid objects

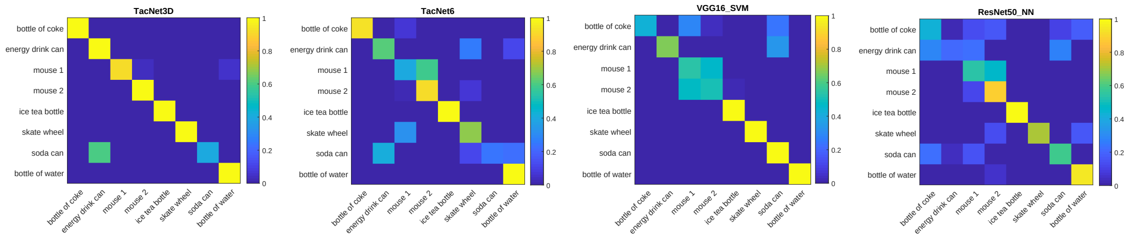
Eight objects of the dataset are considered as rigid because they barely change their shape when the gripper tightens them. The rigid dataset is composed of subsets of objects with similar features (e.g., the subset of bottles and the subset of cans) which are very different from each other. The subset of rigid objects is shown in 5 (a).

4.1.3. Deformable objects

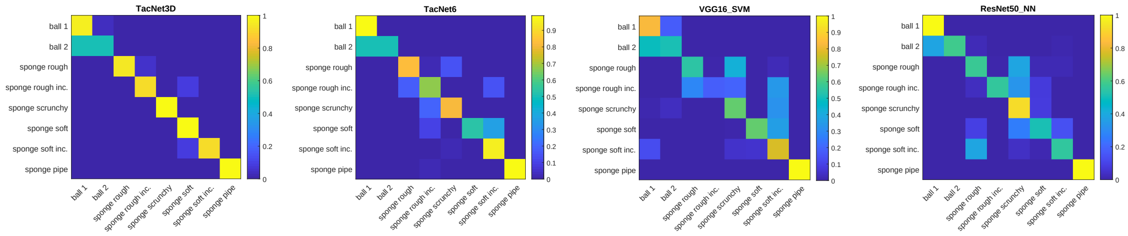
Another subset of the dataset are the deformable objects. This subset consists of eight objects that change substantially his initial shape when a pressure is applied over it, but recover its initial shape when the pressure ends. This subgroup also has objects with similar elasticity (e.g., balls and sponges). The set of deformable objects is shown in 5 (b).

4.1.4. In-bag objects

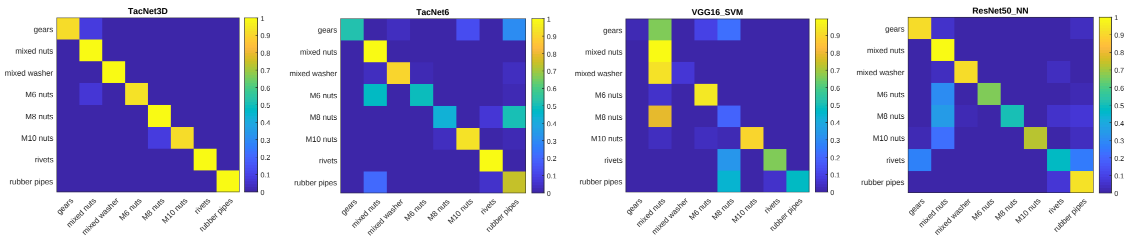
The last subset of objects included in the dataset is composed by plastic bags with a number of small objects. Bags are shuffled before every grasp, so that the objects in the bag are placed in different



(a)



(b)



(c)

Figure 7. Confusion matrices of the methods, from left to right, TactNet3D, TactNet6, VGG16_SVM and ResNet50_NN, in experiments with rigid objects (a), deformable objects (b) and in-bag objects (c). All the methods are trained using data from 2 grasps.

positions and orientations. Hence, the tactile images are different depending on the position of the objects. Another characteristic of this group is that in-bag objects may change their position randomly during the grasping process. As in the other subgroups, bags with similar objects have been chosen (e.g., M6, M8 or M10 nuts). In-bag objects are shown in 5 (c).

4.2. Experiments and results

According to [45], three approaches can be followed to classify tactile data with 2D CNNs: training the network from scratch (method 1), using a pre-trained network with standard images and re-training the last classification layers (method 2), or changing the last layers by other estimator (method 3). The best results for each approach were obtained by TactNet6, ResNet50_NN, and VGG16_SVM, respectively. In this work, four experiments have been carried out to validate and compare the performance of TactNet3D against these 2D CNNs structures considering only the subset of rigid objects, the subset of deformable objects, the subset of in-bag objects, and the whole dataset. The training, validation and test sets to train the 2D CNN-based methods are formed using the individual images extracted from the 3D tactile tensors.

The performance of each method has been measured in terms of recognition accuracy. Each network has been trained 20 times with each subset and the mean recognition rate and standard deviation for each set of 20 samples have been compared in Fig. 6, where for each experiment, the results of each method have been obtained using data from 1, 2, 5, 10 and 20 grasps of each object.

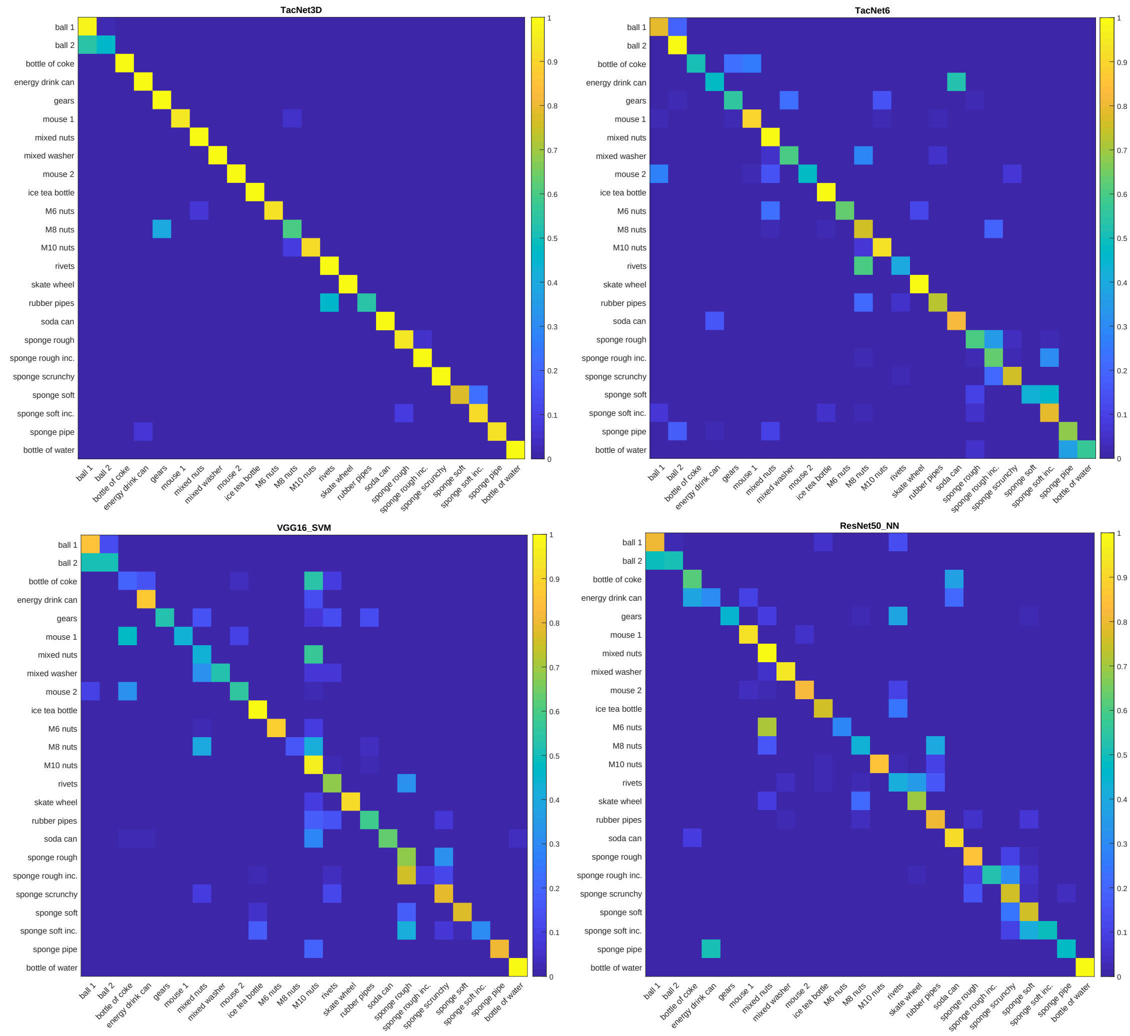


Figure 8. Confusion matrices of the methods, from left to right, TactNet3D, TactNet6, VGG16_SVM and ResNet50_NN, in the experiments with the whole dataset. All the methods are trained using data from 2 grasps.

Moreover, representative confusion matrices for each method trained in subsets of rigid, deformable, and in-bag objects are presented in Fig. 7. In contrast, the confusion matrices related to the whole dataset are presented in Fig. 8. These confusion matrices have been obtained for the case in which each method is trained using data from two grasps to show the differences in classification performance.

5. Discussion

Regarding the performance of TactNet3D in comparison with 2D CNN-based methods, the results shown in 6 prove that the recognition rate of the first one is better than the latter in all the studied cases. For all kinds of objects, rigid, deformable, or in-bag, and all the amount of grasps used as training data, TactNet3D outperforms 2D CNNs.

Also, the differences in classification accuracy are higher when the number of training data is lower, getting better results when training TactNet3D with one or two grasps than 2D CNNs with five or ten grips in some cases. Therefore, it is not only shown that the performance is better, but also the adaptability of TactNet3D as the amount of data needed to train the network is lower, which is especially interesting for online-learning.

Besides, in the misclassification cases, the resulting object class given by TactNet3D has almost indistinguishable physical features to those of the grasped object, unlike 2D CNNs, which may provide disparate results, as can be seen in the confusion matrices presented in Fig. 7 and 8. Looking at some object subsets with similar physical features such as the sponges, the different bag of nuts or the cans, it can be observed that the output given by TactNet3D corresponds to objects form the same subset, whereas 2D CNNs output classes of objects with different features in some cases (e.g., bottle of coke and M10 nuts in Fig. 8 bottom left). This phenomena is interesting from the neurological point of view of artificial touch sense as 3DTactNet behaves more similar to human beings' sense of touch. However, a broad study of this aspect is out of the scope of this paper and will be considered in future works.

6. Conclusions

A novel method for the active tactile perception based on 3D CNN has been presented and used for an object recognition problem in a new robot gripper design. This gripper includes two underactuated fingers that accommodate to the shape of different objects, and have additional proprioceptive sensors to get its actual position. A tactile sensor has been integrated into the gripper, and a novel representation of sequences of tactile images as 3D tactile tensors has been described.

A new 3D CNN has been designed and tested with a set of 24 objects classified in three main categories that include rigid, deformable, and in-bag objects. There are very similar objects in the set, and objects that have changing and complex shapes such as sponges or bags of nuts, to assess the recognition capabilities. 3D CNN and classical CNN with 2D tensors have been tested for comparison. Both perform well with high recognition rates when the amount of training data is high. Nevertheless, 3D CNN gets higher performance even with a lower number of training samples, and misclassification is obtained just in very similar classes.

As future works, we propose the use of additional proprioceptive information to train multi-channel neural networks using the kinesthetic information about the shape of the grasped object, along with the tactile images for multi-modal tactile perception. Also, the use of other dynamic approaches, such as temporal methods (e.g., LSTMs), for both tactile-based and multi-modal-based perception strategies, need to be addressed in more detail. Moreover, a comparison of new active tactile perception methods will be studied in depth.

Author Contributions: conceptualization, F.P. and J.M.G.; software, F.P.; validation, F.P., J.G.G. and J.M.G.; investigation, F.P. J.G.G. and J.M.G.; data curation, F.P.; writing—original draft preparation, F.P., J.G.G. and J.M.G.; writing—review and editing, J.G.G. and J.M.G.; visualization, F.P. and J.M.G.; supervision, J.G.G; project administration, A.G.C. and J.G.G; funding acquisition, A.G.C.

Funding: This research was funded by the University of Málaga, the Ministerio de Ciencia, Innovación y Universidades, Gobierno de España, grants number DPI2015-65186-R and RTI2018-093421-B-I00, and the European Commission, grant number BES-2016-078237.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems* **2012**, pp. 1–9.
2. Cao, L.; Sun, F.; Liu, X.; Huang, W.; Kotagiri, R.; Li, H. End-to-End ConvNet for Tactile Recognition Using Residual Orthogonal Tiling and Pyramid Convolution Ensemble. *Cognitive Computation* **2018**, pp. 1–19.
3. Shibata, A.; Ikegami, A.; Nakauma, M.; Higashimori, M. Convolutional Neural Network based Estimation of Gel-like Food Texture by a Robotic Sensing System. *Robotics* **2017**, *6*.
4. Gandarias, J.M.; Gómez-de Gabriel, J.M.; García-Cerezo, A.J. Tactile Sensing and Machine Learning for Human and Object Recognition in Disaster Scenarios. Third Iberian Robotics conference. Springer, 2017.
5. Vidal-Verdú, F.; Oballe-Peinado, Ó.; Sánchez-Durán, J.A.; Castellanos-Ramos, J.; Navas-González, R. Three realizations and comparison of hardware for piezoresistive tactile sensors. *Sensors* **2011**, *11*, 3249–3266.

6. Chathuranga, D.S.; Wang, Z.; Noh, Y.; Nanayakkara, T.; Hirai, S. Magnetic and Mechanical Modeling of a Soft Three-Axis Force Sensor. *IEEE Sensors Journal* **2016**, *16*, 5298–5307.
7. Ward-Cherrier, B.; Pestell, N.; Cramphorn, L.; Winstone, B.; Giannaccini, M.E.; Rossiter, J.; Lepora, N.F. The TacTip Family: Soft Optical Tactile Sensors with 3D-Printed Biomimetic Morphologies. *Soft Robotics* **2018**, *5*, 216–227.
8. Gong, D.; He, R.; Yu, J.; Zuo, G. A pneumatic tactile sensor for co-operative robots. *Sensors* **2017**, *17*, 2592.
9. Maiolino, P.; Maggiali, M.; Cannata, G.; Metta, G.; Natale, L. A Flexible and Robust Large Scale Capacitive Tactile System for Robots. *IEEE Sensors Journal* **2013**, *13*, 3910–3917.
10. Gandarias, J.M.; Gómez-de Gabriel, J.M.; García-Cerezo, A.J. Enhancing Perception with Tactile Object Recognition in Adaptive Grippers for Human–Robot Interaction. *Sensors* **2018**, *18*, 692.
11. Chitta, S.; Sturm, J.; Piccoli, M.; Burgard, W. Tactile sensing for mobile manipulation. *IEEE Transactions on Robotics* **2011**, *27*, 558–568.
12. James, J.W.; Pestell, N.; Lepora, N.F. Slip Detection With a Biomimetic Tactile Sensor. *IEEE Robotics and Automation Letters* **2018**, *3*, 3340–3346.
13. Romeo, R.; Oddo, C.; Carrozza, M.; Guglielmelli, E.; Zollo, L. Slippage Detection with Piezoresistive Tactile Sensors. *Sensors* **2017**, *17*, 1844.
14. Gandarias, J.M.; Gomez-de Gabriel, J.M.; Garcia-Cerezo, A. Human and object recognition with a high-resolution tactile sensor. IEEE Sensors Conference, 2017.
15. Luo, S.; Mou, W.; Althoefer, K.; Liu, H. Iterative Closest Labeled Point for Tactile Object Shape Recognition. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016.
16. Yuan, Q.; Wang, J. Design and Experiment of the NAO Humanoid Robot's Plantar Tactile Sensor for Surface Classification. 4th International Conference on Information Science and Control Engineering (ICISCE), 2017.
17. Hoelscher, J.; Peters, J.; Hermans, T. Evaluation of tactile feature extraction for interactive object recognition. IEEE-RAS International Conference on Humanoid Robots, 2015.
18. Luo, S.; Bimbo, J.; Dahiya, R.; Liu, H. Robotic tactile perception of object properties: A review. *Mechatronics* **2017**, *48*, 54–67.
19. Trujillo-Leon, A.; Bachta, W.; Vidal-Verdu, F. Tactile Sensor-Based Steering as a Substitute of the Attendant Joystick in Powered Wheelchairs. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2018**, *26*, 1381–1390.
20. Schiefer, M.A.; Graczyk, E.L.; Sidik, S.M.; Tan, D.W.; Tyler, D.J. Artificial tactile and proprioceptive feedback improves performance and confidence on object identification tasks. *PLOS ONE* **2018**, *13*.
21. Bartolozzi, C.; Natale, L.; Nori, F.; Metta, G. Robots with a sense of touch. *Nature Materials* **2016**, *15*, 921–925.
22. Jamone, L.; Natale, L.; Metta, G.; Sandini, G. Highly Sensitive Soft Tactile Sensors for an Anthropomorphic Robotic Hand. *IEEE Sensors Journal* **2015**, *15*, 4226–4233.
23. Roncone, A.; Hoffmann, M.; Pattacini, U.; Fadiga, L.; Metta, G. Peripersonal space and margin of safety around the body: Learning visuo-tactile associations in a humanoid robot with artificial skin. *PLoS ONE* **2016**, *11*, e0163713.
24. Tanaka, Y.; Nagai, T.; Sakaguchi, M.; Fujiwara, M.; Sano, A. Tactile sensing system including bidirectionality and enhancement of haptic perception by tactile feedback to distant part. IEEE World Haptics Conference (WHC), 2013, pp. 145–150.
25. Luo, S.; Mou, W.; Althoefer, K.; Liu, H. Novel Tactile-SIFT Descriptor for Object Shape Recognition. *IEEE Sensors Journal* **2015**, *15*, 5001–5009.
26. Lee, H.; Wallraven, C. Exploiting object constancy: Effects of active exploration and shape morphing on similarity judgments of novel objects. *Experimental brain research* **2013**, *225*, 277–289.
27. Lepora, N.F. Biomimetic Active Touch with Fingertips and Whiskers. *IEEE Transactions on Haptics* **2016**, *9*, 170–183.
28. Okamura, A.M. Feature Detection for Haptic Exploration with Robotic Fingers. *The International Journal of Robotics Research* **2001**, *20*, 925–938.
29. Lepora, N. Active Tactile Perception. In *Scholarpedia of Touch*; Atlantis Press, 2016; pp. 151–159.
30. Dahiya, R.S.; Metta, G.; Valle, M.; Sandini, G. Tactile sensing-from humans to humanoids. *IEEE Transactions on Robotics* **2010**, *26*, 1–20.

31. Zapata-Impata, B.; Gil, P.; Torres, F.; Zapata-Impata, B.S.; Gil, P.; Torres, F. Learning Spatio Temporal Tactile Features with a ConvLSTM for the Direction Of Slip Detection. *Sensors* **2019**, *19*, 523.
32. Drimus, A.; Kootstra, G.; Bilberg, A.; Kragic, D. Design of a flexible tactile sensor for classification of rigid and deformable objects. *Robotics and Autonomous Systems* **2014**, *62*, 3–15.
33. Dolz, J.; Desrosiers, C.; Ayed, I.B. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage* **2018**, *170*, 456–470.
34. Chaddad, A.; Desrosiers, C.; Niazi, T. Deep radiomic analysis of MRI related to Alzheimer's Disease. *IEEE Access* **2018**, *6*, 58213–58221.
35. Gandarias, J.M.; Pastor, F.; García-Cerezo, A.J.; Gómez-de Gabriel, J.M. Active Tactile Recognition of Deformable Objects with 3D Convolutional Neural Networks. IEEE World Haptics Conference (WHC), 2019, pp. 551–555.
36. Feng, D.; Kaboli, M.; Cheng, G. Active Prior Tactile Knowledge Transfer for Learning Tactual Properties of New Objects. *Sensors* **2018**, *18*, 634.
37. Kaboli, M.; Cheng, G. Robust Tactile Descriptors for Discriminating Objects From Textural Properties via Artificial Robotic Skin. *IEEE Transactions on Robotics* **2018**, pp. 1–19.
38. Baishya, S.S.; Bauml, B. Robust material classification with a tactile skin using deep learning. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016.
39. Jamali, N.; Sammut, C. Majority voting: Material classification by tactile sensing using surface texture. *IEEE Transactions on Robotics* **2011**, *27*, 508–521.
40. Liu, H.; Song, X.; Nanayakkara, T.; Seneviratne, L.D.; Althoefer, K. A computationally fast algorithm for local contact shape and pose classification using a tactile array sensor. IEEE International Conference on Robotics and Automation (ICRA), 2012.
41. Martinez-Hernandez, U.; Dodd, T.J.; Prescott, T.J. Feeling the Shape: Active Exploration Behaviors for Object Recognition With a Robotic Hand. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **2017**, pp. 1–10.
42. Yi, Z.; Calandra, R.; Veiga, F.; van Hoof, H.; Hermans, T.; Zhang, Y.; Peters, J. Active tactile object exploration with gaussian processes. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016.
43. Corradi, T.; Hall, P.; Iravani, P. Bayesian tactile object recognition: Learning and recognising objects using a new inexpensive tactile sensor. IEEE International Conference on Robotics and Automation (ICRA), 2015.
44. Albini, A.; Denei, S.; Cannata, G. Human Hand Recognition From Robotic Skin Measurements in Human-Robot Physical Interactions. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017.
45. Gandarias, J.M.; García-Cerezo, A.J.; Gómez-de Gabriel, J.M. CNN-based Methods for Object Recognition with High-Resolution Tactile Sensors. *IEEE Sensors Journal* **2019**.
46. Falco, P.; Lu, S.; Cirillo, A.; Natale, C.; Pirozzi, S.; Lee, D. Cross-modal visuo-tactile object recognition using robotic active exploration. IEEE International Conference on Robotics and Automation (ICRA), 2017.
47. Luo, S.; Liu, X.; Althoefer, K.; Liu, H. Tactile object recognition with semi-supervised learning. International Conference on Intelligent Robotics and Applications, 2015.
48. Khasnobish, A.; Jati, A.; Singh, G.; Bhattacharyya, S.; Konar, A.; Tibarewala, D.; Kim, E.; Nagar, A.K. Object-shape recognition from tactile images using a feed-forward neural network. International Joint Conference on Neural Networks (IJCNN), 2012.
49. Schmitz, A.; Bansho, Y.; Noda, K.; Iwata, H.; Ogata, T.; Sugano, S. Tactile object recognition using deep learning and dropout. IEEE-RAS International Conference on Humanoid Robots, 2014.
50. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Transactions on Neural Networks* **1997**, *8*, 98–113.
51. Albini, A.; Denei, S.; Cannata, G. Human hand recognition from robotic skin measurements in human-robot physical interactions. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017, pp. 4348–4353.
52. Madry, M.; Bo, L.; Kragic, D.; Fox, D. ST-HMP: Unsupervised Spatio-Temporal feature learning for tactile data. IEEE International Conference on Robotics and Automation (ICRA), 2014.
53. Liu, H.; Guo, D.; Sun, F. Object Recognition Using Tactile Measurements: Kernel Sparse Coding Methods. *IEEE Transactions on Instrumentation and Measurement* **2016**, *65*, 656–665.

54. Kerzel, M.; Ali, M.; Ng, H.G.; Wermter, S. Haptic material classification with a multi-channel neural network. International Joint Conference on Neural Networks (IJCNN), 2017.
55. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3d convolutional networks. Proceedings of the IEEE international conference on computer vision, 2015, pp. 4489–4497.
56. Birglen, L.; Laliberté, T.; Gosselin, C.M. *Underactuated robotic hands*; Vol. 40, Springer, 2007.