Active Tactile Recognition of Deformable Objects with 3D Convolutional Neural Networks

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

Abstract— In this paper, a new concept of active tactile perception based on deep learning is presented. A tactile sensor is used to acquire sequences of tactile images of deformable objects when different forces are applied. Hence, the sequence of data can be represented by 3D tactile tensors in a similar way to the sequences of images represented in Magnetic Resonance Imaging (MRI). However, in this case, each 2D frame represents the pressure distribution when a certain force is applied, and the third dimension represents time or the variation of the applied force. Due to this feature of data, a 3D Convolutional Neural Network (3D CNN) called TactNet3D has been created to classify tactile information from 9 deformable objects. A dataset composed of 540 tactile sequences formed by [28x50x10] tactile tensors is used to train, validate and test the performance of TactNet3D, showing that it can classify deformable objects with an accuracy of 96.39% with time series of pressure distributions.

I. Introduction

Artificial tactile perception can be defined as the integration of tactile sensing and artificial intelligence to obtain high-level information. This kind of perception is as critical for robotics manipulation as the sense of touch is for humans [1]. In fact, tactile perception has become one of the most important research topics in robotics, and it has been demonstrated that this information is extremely useful for robotic systems [2], [3], [4]. The number of applications that can benefit from it is huge and is not limited to robotics (e.g. medicine [5], food industry [6] or search and rescue [7] among others). Existing studies are focused on classifying materials or objects by their shape, stiffness and roughness [8], [9].

Multiple tactile object recognition related works are based on machine learning algorithms. In a recent study, an autonomous robotic palpation system uses a machine learning algorithm to locate and segment hard inclusions in artificial tissues [10]. To apply machine learning, existing solutions treat tactile data as images. That way, in [11] a novel algorithm synthesizes kinesthetic and tactile images to form a 4D point cloud to identify touched objects. Following this approach, Deep Learning-based methods have also been used to classify tactile array data by treating the sensed information as images. In [12] a tactile sensor is attached to
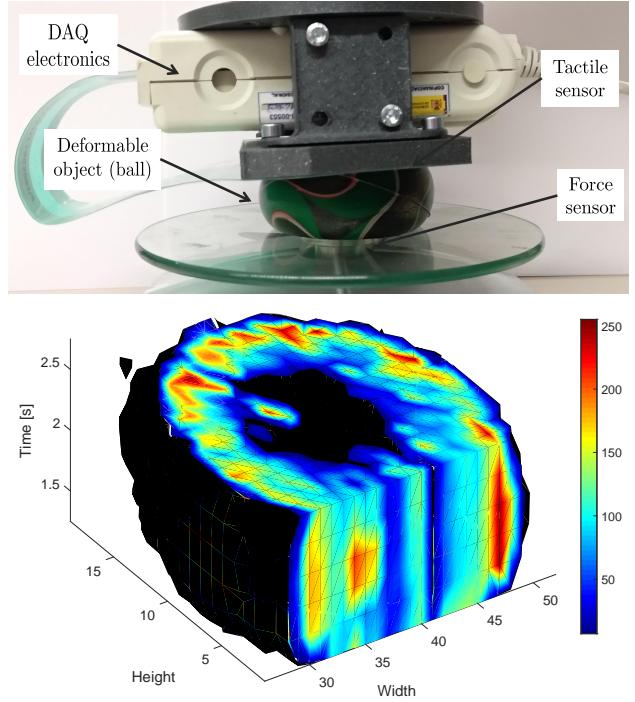
Fig. 1. A tactile sensor is used to obtain sequences of tactile images from a deformable object (top). The pressure distribution is acquired when different forces are applied and the sequences of pressure distributions can be represented through time (bottom).

an adaptive robotic gripper to classify in-hand objects with Convolutional Neural Networks (CNNs), and in [13], a CNN is used to classify in-contact human hand with an artificial skin. In [14], multiple CNN-based methods to recognize non-deformable objects with high-resolution tactile sensors are presented.

Although existing methods have obtained good results in tactile object recognition, one of the main drawbacks is that most data acquisition processes might be considered passive, as the information are static tactile images [15], which is not a natural haptic exploratory procedure to perceive pressure, stiffness or shape information from a touched object [16]. Static pressure images only have information about the shape or stiffness of in-contact objects when a certain force is applied [17], but a sequence of tactile images has information about the variation of shape, pressure and stiffness of an object when different forces are applied [18]. This kind of information is specially useful when dealing with

deformable objects [19]. To classify sequences of tactile data, a simple Two-Dimensional Convolutional Neural Network (2D CNN) can be used, in [6], pressure data is acquired by a flexible tactile array and a 2D CNN is used to classify a sequence of pressure images. However, as it is demonstrated in [20], Three-Dimensional Convolutional Neural Networks (3D CNNs) can classify sequences of images better than common 2D CNNs.

In this paper, a novel artificial tactile perception approach based on an active haptic exploration procedure is presented and validated. A tactile sensor has been used to obtain sequences of tactile images (See Fig. 1) from deformable objects in an active data acquisition procedure. Tactile data is represented as 3D tensors in a similar way to Magnetic Resonance Imaging (MRI), which contains information of cross-sectional images of internal organs and structures over distance or time, whereas 3D tactile tensors contain information of the variation of pressure distributions over time. The main contribution of this paper is to propose a novel 3D CNN-based method for active tactile perception. In order to show the performance of the method, a 3D CNN called TacNet3D is designed to classify active tactile data in an application to recognize deformable objects. Although the experiment presented in this paper, which is only used to validate the proposed methodology, does not have a direct application, this approach can be applied to multiple fields such as touch-based object classification, human-robot interaction, robotic manipulation or food and medical industries, among others. The experiment consists of classifying 9 deformable objects, which has not been chosen only to show that the proposed method is able to accurately recognize objects with similar shape and stiffness, but also identical objects with or without hard inclusions, what could not be done with common static-based tactile perception methodologies.

This paper is structured as follows: in section II the 3D CNN-based method used for tactile data classification is presented, in section III the experimental procedure, sensor and data specifications and results are described. Finally, in IV the conclusions of this work and future research lines are exposed.

II. METHODOLOGY

This section explains how 3D tactile information is represented, and describes the 3D CNN designed for this work.

A. Representation of 3D tactile information

The natural exploratory procedure to get information about the stiffness is to palpate an object and perceive the information during the whole squeeze-and-release process. The representation of this information when the process is carried out using an artificial tactile sensor can be represented as a set or sequence of tactile images. Here, all the frames of each sequence are used to form 3D tactile tensors. This representation process imitates the representation of MRI data, but in this case each frame contains information of pressure distribution over time instead of cross-sectional images of internal organs and structures. In Fig. 2, 3D
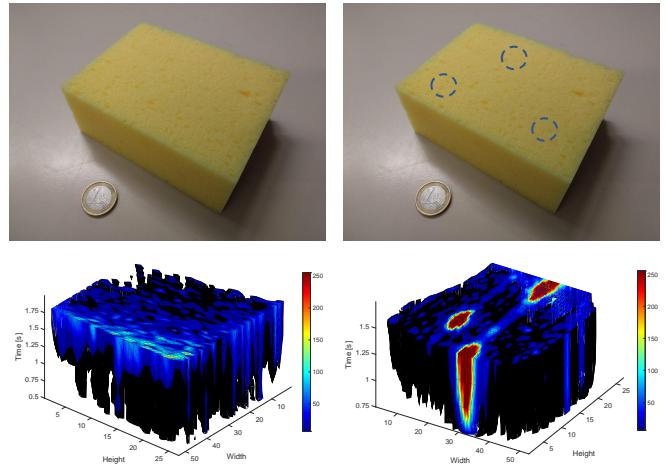

Fig. 2. Representation of 3D tactile tensors of a sponge (left) and the same sponge with hard inclusions (right). 3D images are sectioned to show the cross-sectional pressure distribution. Note that since inclusions are inside the object and cannot be seen in a picture, they are represented with the blue circles.

tactile tensors of an object with and without hard inclusions is shown. The process followed to collect tactile data is described in section III.

Although MRI and 3D tactile information are different, they can be represented in the same way, therefore, methods used to process MRI information might be used for tactile data. In [21], [22], 3D CNNs have been used to process MRI data with satisfactory results.

B. TactNet3D

In regards to 3D tactile processing, the point of 3D CNNs in comparison to 2D CNNs is that in the second case a pressure control is needed to get the most representative tactile image of an object, whereas in the first case, as 3D tactile tensors contain information of the pressure distributions over the whole palpation process, this pressure control is unnecessary. This aspect is even more remarkable when dealing with deformable objects, as 2D CNNs need optimal tactile images to characterize objects, and it is hard to find this optimal pressure distributions because they depend on the properties of the objects themselves.

Designing a 3D CNN that performs well is difficult as a network could have multiple architectures and parameters. Therefore, to find a good network design, four 3D CNNs have been designed and evaluated. The results of this evaluation are commented in section III and summarized in Table I. The best configuration is a 4-layer 3D CNN, called TacNet3D, that has been created to classify deformable objects from 3D tactile tensors. The architecture of TactNet3D is presented in Fig. 3.

TactNet3D is composed by two 3D convolutional layers ($\mathscr{C} = [\text{3D conv}_1, \text{3D conv}_2]$) with kernels $8 \times [5 \times 5 \times 3]$ and $16 \times [5 \times 3 \times 3]$ respectively, and two fully connected layers ($\mathscr{F} = [\text{fc}_3, \text{fc}_4]$) with 64 and 9 neurons respectively. Each 3D convolutional layer includes batch normalization with $\varepsilon =$
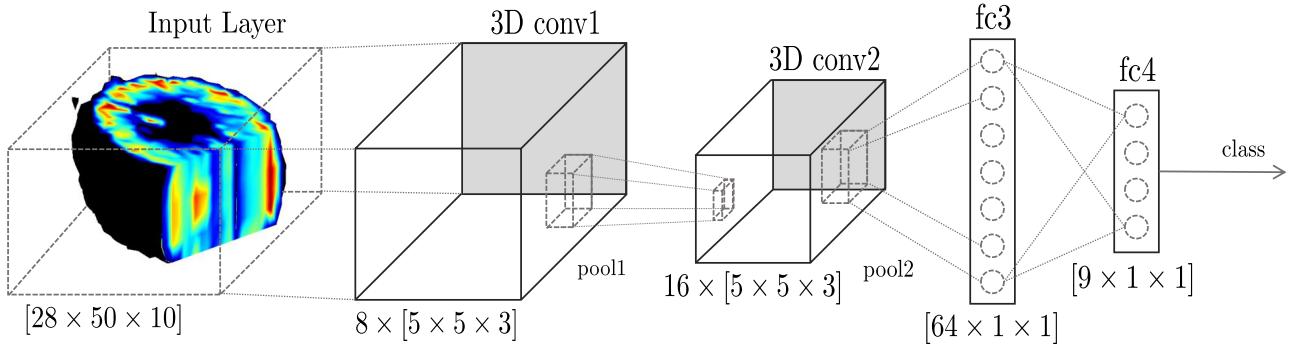
Fig. 3. Architecture of the 4-layers TactNet3D.

10^{-5}, a Rectified Linear Unit (ReLU) and max-pooling with filters and stride $= 1$. Each fully-connected layer includes a dropout of 0.8 to prevent overfitting, and the last fully-connected layer is followed by a softmax function that gives the probability of it belonging to each class. The implementation of the 3D CNNs configurations has been carried out using the *mdCNN Matlab Toolbox* [1] running in an Intel Core i7-4790 CPU @ 3.60 GHz with 16 GB of RAM with a random weights initialization. The use of GPUs with this toolbox is not supported so far.

III. EXPERIMENTS

In this section, the sensor specifications, the data collection process, the dataset and the results obtained in the experiments are described.

A. Sensor specifications

To collect the tactile data, the sensor model 6077 provided by Tekscan (South Boston, MA, USA) is used. The tactile sensor has 1400 pressure tactels or taxels distributed in a 28×50 matrix. The size of the matrix is 53.3×95.3 mm, forming a set of resistive pressure sensors with density 27.6 tactels/cm^2 and maximum pressure admitted of 34 kPa. The system integrates the data acquisition electronics (DAQ) called *Evolution Handle*, and the *I-Scan* software, which is used to supervise the data collection process.

B. Data collection process

Data collection is a key task in machine learning, therefore, data has been collected manually and supervised by an operator to ensure that the dataset is captured correctly. Finally, a dataset formed by 540 3D tactile tensors has been collected, following a process which consists of applying variations of pressures and saving the 3D tactile tensors from each object. This way, 60 sequences of 10 tactile images have been collected for each object. A force sensor situated under the object (Fig. 1) is used to control that the applied force is within the interval $[0, 40]$ N. When different forces are applied different pressure images are collected, depending on the intrinsic features of the contact object. A sequence of matrices of a 3D Tactile tensor is shown in Fig.4, which

[1]https://github.com/hagaygarty/mdCNN

TABLE I
MEAN ACCURACY (Ā), STANDARD DEVIATION (σ) AND TRAINING TIME (t_{training}) OF THE 3D CNNs ARCHITECTURES TESTED

Architecture	Ā [%]	σ [%]	t_{training} [s]
$2 \times$ 3D conv [4,8] $+ 2\times$ fc [64,9]	80.79	4.17	63.19
$2 \times$ 3D conv [8,16] $+ 2\times$ [64,9]	96.39	1.97	155.99
$2 \times$ 3D conv [8,16] $+ 2\times$ [32,9]	91.20	3.13	145.58
$3 \times$ 3D conv [8,16,32] $+ 2\times$ [64,9]	93.29	3.62	600.68

shows 6 different photographs(top) and frames (bottom) of the data collection process of a deformable ball.

For Deep learning-based methods the more amount and variation of data, the better. Even though the applied force is limited in order for it not to exceed 40N, the applied pressure has not been controlled, which means that the force increment between two frames is not constant and may vary for each pair of neighbour frames. This way, the resultant method has to be robust enough to classify deformable objects without the necessity to integrate a pressure control method.

C. Dataset

A dataset of 540 3D tactile tensors from 9 objects is collected to train, validate and test the performance of the TactNet3D. Each data sample consists of a 3D Tactile tensor of 10 pressure distributions, and there are 60 samples of each object. The deformable objects shown in Fig. 5, are labeled as: *ball_inclusions, ball_rough, ball, sponge_exfoliating, sponge_smooth, sponge_smooth_inclusions, sponge_scrunchy, sponge_rough* and *sponge_rough_inclusions*.

To train, validate and test the 3D CNN, the dataset has been split into training, validation and test sets, which respectively have 432 (80%), 54 (10%) and 108 (20%) tactile tensors of the original dataset formed by 540 3D tactile tensors.

D. Results

As commented in section II, four 3D CNN configurations have been evaluated, and the best result (TactNet3D) has two
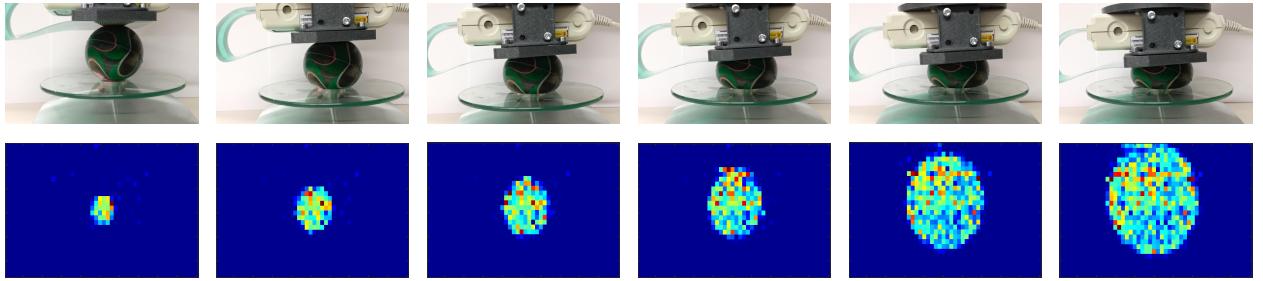
Fig. 4.   Illustrations of the experimental process (top) followed to collect sequences of tactile images (bottom) from a deformable ball.

Fig. 5.   Deformable objects used in experiments. From top left to bottom right: *ball_inclusions, ball_rough, ball, sponge_exfoliating, sponge_smooth, sponge_smooth_inclusions, sponge_scrunchy, sponge_rough* and *sponge_rough_inclusions*. Note that since inclusions are inside the object and cannot be seen in a picture, they are represented with blue circles.

3D convolutional layers and two fully connected layers. To evaluate their performance, each network was been tested 20 times with random training, validation and test sets for each sample. Table I presents the architectures of each network and summarizes the results in terms of mean accuracy ($\bar{A}$), standard deviation σ and training time $t_{training}$. TactNet3D is represented by the second architecture in the table, and achieves the highest mean accuracy (96.39%) and the lowest standard deviation (1.97%) which demonstrate the good performance and robustness of the network. However, the lowest training time (63.19 s) is achieved by the first configuration as it is the shallowest network. The deepest configuration tested is the last configuration in the table, and it has three 3D convolutional layers and two fully connected layers. It can be seen that despite being the deepest network, as the dataset is small, the results of this model are not the best.

The training process of the network has also been monitored. The success rate and loss achieved by training and validation sets during the training process of TactNet3D are shown in Fig. 6. The loss is computed using the cross entropy loss function.

Finally, the evaluation of TactNet3D in the 9-classes classification problem can be discussed according to the confusion matrix obtained in a classification experiment (see Fig. 7).

It must be said that these results may vary depending on the data used for training, validation and test sets, which are chosen randomly. In any case, this figure shows the good performance of the network in this case, achieving a
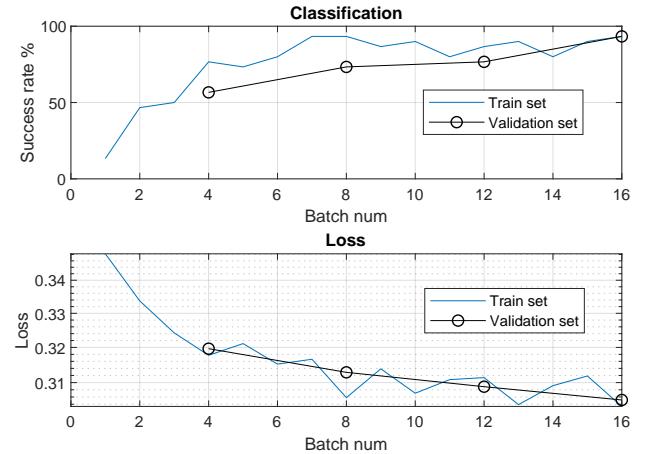

Fig. 6.   Success rate (top) and loss (bottom) during the training process of the TactNet3D.

mean accuracy of 96.30%. With these sets of data TactNet3D achieve the 100% of accuracy in all of the objects except the object 1 - 75% (*ball_inclusions*) and the object 7 - 91.67% (*sponge_scrunchy*).

IV. CONCLUSIONS

The problem of artificial tactile recognition of deformable objects has been addressed in this paper. This problem has been faced considering that, since natural tactile perception is active, artificial tactile perception should also be active.
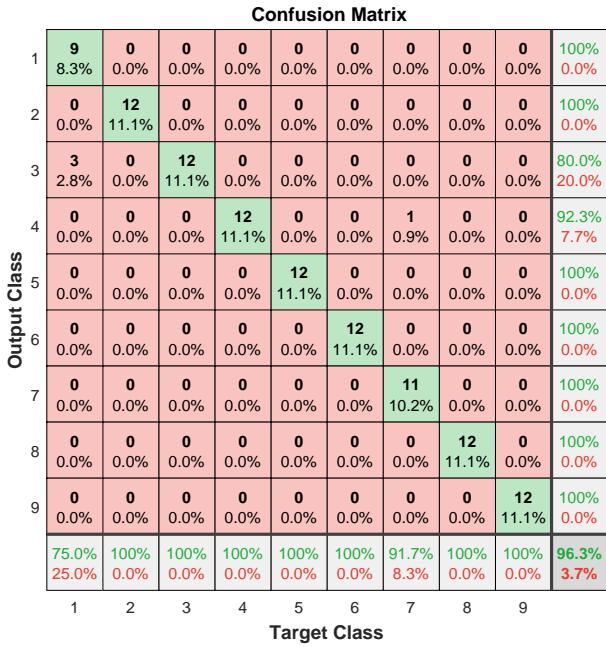
Fig. 7. Confusion matrix for a random training, validation and test sets of TactNet3D. Classes are enumerated from 1 to 9, rows represent the prediction of the network (Output Class) and columns are the actual class (Target Class). Diagonal cells represents the percentage of correct classifications for each class.

Therefore, 3D tactile data have been collected and represented as 3D tactile tensors. These tensors do not only have information about pressure distribution, but also their variation over time when different pressures are applied during a palpation exploratory procedure of squeezing and releasing the contact object. A 3D CNN, called TactNet3D, has been designed and validated in a 9-classes classification problem with deformable objects. Results have shown the good performance of the network in the classification task (96.39%) although the dataset is small (540 tactile tensors).

Although the performance of the method has been validated in an experiment with deformable objects only, its expected to present a good performance with non-deformable objects according to results of previous works that used 3D CNNs for computer vision applications. However, this statement needs to be demonstrated through experimentation in future works. Moreover, an experiment with more classes, mixing rigid and deformable objects, and data augmentation techniques to enlarge the dataset will be considered. Furthermore, the application of these techniques to robotic manipulation will be studied, integrating the presented approach in robotic grippers for practical applications. Finally, the performance of 3D CNNs has to be compared against other active and passive tactile methods such as 2D CNNs or CNN-RNN models.

REFERENCES

[1] M. A. Schiefer, E. L. Graczyk, S. M. Sidik, D. W. Tan, and D. J. Tyler, "Artificial tactile and proprioceptive feedback improves performance and confidence on object identification tasks," *PLOS ONE*, vol. 13, no. 12, 2018.

[2] A. Trujillo-Leon, W. Bachta, and F. Vidal-Verdu, "Tactile sensor-based steering as a substitute of the attendant joystick in powered wheelchairs," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 7, pp. 1381–1390, 2018.

[3] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, "Single-grasp object classification and feature extraction with simple robot hands and tactile sensors," *IEEE Transactions on Haptics*, vol. 9, no. 2, pp. 207–220, 2016.

[4] S. E. Navarro, N. Gorges, H. Wörn, J. Schill, T. Asfour, and R. Dillmann, "Haptic object recognition for multi-fingered robot hands," in *Haptics Symposium*, 2012.

[5] Y. Tanaka, T. Nagai, M. Sakaguchi, M. Fujiwara, and A. Sano, "Tactile sensing system including bidirectionality and enhancement of haptic perception by tactile feedback to distant part," in *IEEE World Haptics Conference (WHC)*, 2013, pp. 145–150.

[6] A. Shibata, A. Ikegami, M. Nakauma, and M. Higashimori, "Convolutional Neural Network based Estimation of Gel-like Food Texture by a Robotic Sensing System," *Robotics*, vol. 6, no. 4, 2017.

[7] J. M. Gandarias, J. M. Gómez-de Gabriel, and A. J. García-Cerezo, "Tactile sensing and machine learning for human and object recognition in disaster scenarios," in *Third Iberian Robotics conference*. Springer, 2017.

[8] Q. Yuan and J. Wang, "Design and Experiment of the NAO Humanoid Robot's Plantar Tactile Sensor for Surface Classification," in *4th International Conference on Information Science and Control Engineering (ICISCE)*, 2017.

[9] U. Martinez-Hernandez, T. J. Dodd, L. Natale, G. Metta, T. J. Prescott, and N. F. Lepora, "Active contour following to explore object shape with robot touch," in *IEEE World Haptics Conference (WHC)*, 2013, pp. 341–346.

[10] K. A. Nichols and A. M. Okamura, "Methods to Segment Hard Inclusions in Soft Tissue During Autonomous Robotic Palpation," *IEEE Transactions on Robotics*, vol. 31, no. 2, pp. 344–354, 2015.

[11] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Iterative Closest Labeled Point for Tactile Object Shape Recognition," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[12] J. M. Gandarias, J. M. Gómez-de Gabriel, and A. J. García-Cerezo, "Enhancing Perception with Tactile Object Recognition in Adaptive Grippers for HumanRobot Interaction," *Sensors*, vol. 18, no. 3, p. 692, 2018.

[13] A. Albini, S. Denei, and G. Cannata, "Human Hand Recognition From Robotic Skin Measurements in Human-Robot Physical Interactions," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.

[14] J. M. Gandarias, A. J. García-Cerezo, and J. M. Gómez-de Gabriel, "CNN-based methods for object recognition with high-resolution tactile sensors," *IEEE Sensors Journal*, 2019.

[15] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Novel Tactile-SIFT Descriptor for Object Shape Recognition," *IEEE Sensors Journal*, vol. 15, no. 9, pp. 5001–5009, 2015.

[16] N. F. Lepora, "Biomimetic Active Touch with Fingertips and Whiskers," *IEEE Transactions on Haptics*, vol. 9, no. 2, pp. 170–183, 2016.

[17] J. M. Gandarias, J. M. Gomez-de Gabriel, and A. Garcia-Cerezo, "Human and object recognition with a high-resolution tactile sensor," in *IEEE Sensors Conference*, 2017.

[18] B. Zapata-Impata, P. Gil, F. Torres, B. S. Zapata-Impata, P. Gil, and F. Torres, "Learning Spatio Temporal Tactile Features with a ConvLSTM for the Direction Of Slip Detection," *Sensors*, vol. 19, no. 3, p. 523, 2019.

[19] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Design of a flexible tactile sensor for classification of rigid and deformable objects," *Robotics and Autonomous Systems*, vol. 62, no. 1, pp. 3–15, 2014.

[20] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[21] J. Dolz, C. Desrosiers, and I. B. Ayed, "3d fully convolutional networks for subcortical segmentation in mri: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, 2018.

[22] A. Chaddad, C. Desrosiers, and T. Niazi, "Deep radiomic analysis of mri related to alzheimers disease," *IEEE Access*, vol. 6, pp. 58 213–58 221, 2018.