2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
September 24–28, 2017, Vancouver, BC, Canada

Human Hand Recognition From Robotic Skin Measurements in Human-Robot Physical Interactions

Alessandro Albini, Simone Denei[1] and Giorgio Cannata

***Abstract*— This paper deals with the problem of using the tactile feedback generated by a robotic skin for discriminating a human hand touch from a generic contact. Humans understand collaboration intentions through different sensing modalities such as vision, hearing and touch. Among them, a physical interaction is mainly used for demonstrating or correcting a kind of motion and is usually started by touching with the hands the other human body. Until recently, it was difficult to perform the same in human-robot cooperation due to the lack of large-scale tactile systems functionally similar to a human skin. Our approach consists in transforming measurements of sensors distributed on the robot body into a convenient 2D representation of the contact shape, i.e., a contact image, then applying image classification techniques in order to discriminate a human touch from unexpected collisions. Experiments have been performed on a robotic skin composed of 768 pressure sensors integrated on a Baxter robot forearm. More than 1800 contact images have been generated from 43 different persons for training and testing two machine learning algorithms: Bag of Visual Words and Convolutional Neural Networks. The experimental results show that both approaches are valid, obtaining a classification accuracy higher than 96%.**

I. Introduction

One of the main goals of human-robot cooperation (HRC) is to exploit the capabilities of both in performing assembly tasks [1]. In fact, robots have great ability and accuracy in repetitive operations but their performances decrease when tasks are executed in unstructured environments or are subject to unexpected changes. In such cases, the intervention of a human can improve the overall efficiency.

Humans understand collaboration intentions through multiple sensing modalities such as vision, hearing and touch. Most of the proposed approaches in HRC use computer vision in order to understand actions or intentions of the nearby humans. For example in [2], a robot assists a human for assembling a toy while, in [3], the authors developed a vision based system for collisions detection. Other approaches are based on the robot joint torque measurements, e.g. in [4], where this information is used for guaranteeing a safe HRC. A different approach to HRC exploits an acoustic system in [5].

All the previous approaches have a number of limitations. Indeed, vision fails when an occlusion occurs, while using methods based on torque measurements do not allow to localize the contact on the robot body. The same considerations apply to techniques based on acoustic sensors. It is evident that, if the robot is equipped with a tactile system functionally similar to the human skin, human-robot cooperation can be improved. The recent progress in distributed tactile sensors has enabled the realization of systems composed of different transducers (e.g. pressure, temperature or vibration), that can cover the robot body providing a tactile feedback from a large area [6], [7]. This opens new scenarios on human robot interaction and physical contacts processing where the robot can interpret the human intentions and react accordingly.

Understanding the type of interaction a human can have with a robot is a complex task [8].However, it can be assumed that if a human wants to interact with the robot, e.g., for adjusting its pose or to stop the motion, the most natural way would be by using his hands to grasp a robot limb. For this reason, this work focuses on *the identification of a human hand touch in contact with the robot body through robotic skin measurements*. The aim is to *discriminate a voluntary contact performed by a human from unexpected collisions*. Moreover, what is proposed in this paper could be relevant for the paradigm of *programming by demonstration* where a human teaches a movement to the robot through physical interactions [9].

In literature, tactile sensors have been already used for recognizing objects by processing *contact images*, i.e., a graphical representation of contacts where each pixel value is related to the force/pressure measurements of the tactile system. In [10], a neural network for objects classification has been trained using contact images generated by planar patches of pressure sensors integrated on a robot hand. Using the same concept, the classification has been performed using a Bag of Visual Words model and tactile-SIFT descriptors [11]. A similar approach is presented in [12] where authors use contact images obtained by haptic exploration of objects to be classified.

The approach presented in this paper exploits the same concept of contact image for detecting the touch generated by human hands on the robot body. The main difference with respect to previous works is that pressure sensors are distributed in a nonuniform way over a *3-dimensional surface* (i.e., the robot body). The proposed technique is based on a transformation of tactile data from a 3D manifold into a flat 2D domain, i.e., a contact image. The major advantage of using this approach is that human hand detection can be supported by algorithms already developed for image processing and classification.

Once the contact image is obtained, it can be processed in order to discriminate generic contacts from the human

All the authors are with the Department of Informatics, Bioengineering, Robotics and Systems Engineering (DIBRIS), University of Genoa, Via Opera Pia 13, 16145, Genoa, Italy.
[1]Corresponding author's e-mail: simone.denei@unige.it

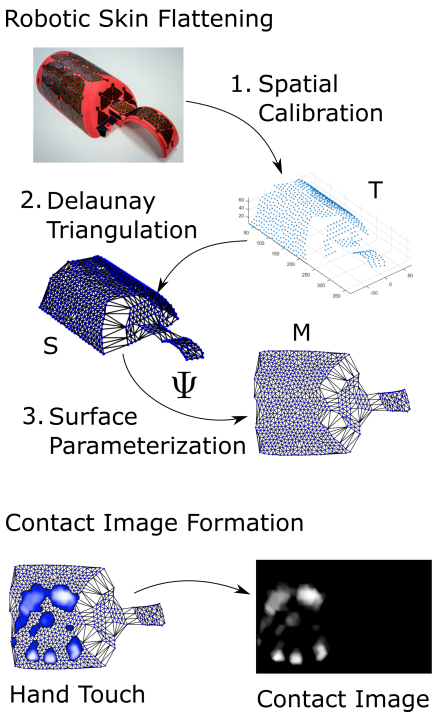
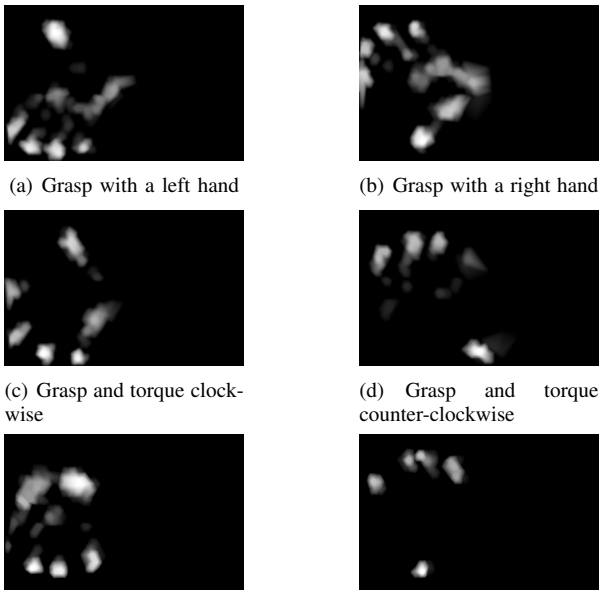


Fig. 1. Overview of the proposed approach.



(a) Grasp with a left hand

(b) Grasp with a right hand

(c) Grasp and torque clockwise

(d) Grasp and torque counter-clockwise

(e) Pushing with a left hand

(f) Pulling with a right hand

Fig. 2. Examples of tactile images generated by a human subject during different interactions with the robot.

hand touch. However, as it can be seen in Figure 2, the hand touch produces different signs on the robotic skin involving the whole hand, part of it or just its fingertips. This mainly depends on the following factors:

- The shape of the body part grasped by the human.
- The human physical characteristics. This includes not only the shape of the hand but also the physical strength, the weight, height and gender just to name but a few.
- The kind of interaction. For example, while pushing away the robot arm requires the whole hand, pulling the same part mainly involves the fingertips.

The variability of the signs produced by a hand contact suggests that machine learning algorithms should be used for human hand contact recognition. In this work, two methods will be investigated that are Bag of Visual Word (BoVW) model and Convolutional Neural Networks (CNNs). Both are widely applied in image classification [13][14] and BoVW has been already used for tactile-based object classification [11]. Moreover, BoVW model has been employed in [15] for classifying hand gestures in real time, as well as CNNs [16], [17], [18]. In addition, CNNs are robust against image variations [19], such as scale and rotation, suggesting their usage for human hand contact classification.

The paper is organized as follows. Section II describes how to obtain a tactile image out of a robotic skin feedback. In Section III different classification algorithms will be defined and used for hand touch detection. Section IV reports the experimental results. Conclusion follows.

II. Contact Image Formation from Distributed Tactile Sensors Measurements

A. The Robotic Skin

Fig. 3. The Baxter forearm covered by 768 tactile elements. It can be seen that the spatial resolution is not uniform and the taxels are not arranged in a regular grid.

The proposed approach has been tested on a prototype of robotic skin similar to the one presented in [6]. It is made by equilateral triangular modules, of $3\,cm$ edge, interconnected to each other forming patches that can be used for covering the robot body. Each module consists of 11 pressure sensors, of $3.5\,mm$ of diameter and $8\,mm$ pitch, and are realized with a flexible printed circuit board that can be conformed to curved surfaces. In Figure 3, the robotic skin partially covers in a non-uniform way the forearm of a Baxter robot

[20]. The position of each sensor with respect to a robot reference frame is supposed to be known as an outcome of a spatial calibration procedure [21], [22]. The result is a set $T = \{\mathbf{t}_1, \ldots, \mathbf{t}_N\}$ where $\mathbf{t}_i \in \Re^3$ is the 3D location of i-th tactile element, i.e. a *taxel*, and N is the number of sensors composing the skin (see Figure 1).

B. Surface Parametrization

Starting from the set of points T, a triangular mesh S approximating the robot body surface is generated by means of *Delalunay triangulation* [23] (see Figure 1). Following the procedure reported in [24], the idea is to convert the triangular mesh in a 2D map that represents the sensorized area. To this aim, the key idea is to use the *Surface Parametrization theory* [25] that allows to topographically preserve sensor locations, displacements, density and proximity relationships among nearby sensors [24]. More formally, it allows to define a piecewise linear mapping $\Psi : S \rightarrow M$ between S and an isomorphic discrete 2D surface $M = \{\mathbf{m}_1, \ldots, \mathbf{m}_N\}$ whose vertices $\mathbf{m}_i \in \Re^2$ best preserve the intrinsic properties of S, i.e. for each $\mathbf{t}_i \in S$, a corresponding $\mathbf{m}_i \in M$ exists such that $\mathbf{t}_i = \Psi^{-1}(\mathbf{m}_i)$. Figure 1 shows the result of the method applied to a set of points that represent the 3D sensor positions on the robot forearm. It can be seen, that the properties cited above are preserved and the introduced distortions are minimal.

C. Contact image formation

During contact, the robotic skin captures the exerted pressure on the robot body by generating a set of pressure measurements $P = \{p_1, p_2, \ldots, p_N\}$, where $p_i \in \Re$ is the measurement of the i-th taxel. While the sets S and P represent the discrete pressure distribution of the contact on the robot body, M and P can be used to generate the related pressure map (see Figure 1). The following step consists in transforming this map that into an *image*, i.e., a regular grid of pixels. The image is obtained as follows. The area of the tactile map M is re-sampled with a regular grid with R rows and C columns where $\mathbf{x}_{r,c}$ is the position of the point in the grid at row r and column c (see Figure 4). The values of R and C are computed in order to obtain a grid spacing comparable to the spatial resolution of the robotic skin (e.g., $8mm$ in our case) in the area covered by M. For each point $\mathbf{x}_{r,c}$, the related pressure value $p_{r,c}$ is computed by means of the *barycentric interpolation* of the nearby taxels. In particular, for a given point $\mathbf{x}_{r,c}$, the pressure value can be computed by finding the enclosing triangle defined by taxel positions $\mathbf{m}_j, \mathbf{m}_k$ and $\mathbf{m}_h$ and their pressure measurements p_j, p_k and p_h (see Figure 4), and by using the following equation:

$$p_{r,c} = \frac{(A_1 p_h + A_2 p_k + A_3 p_j)}{A},$$

where A is the area of the triangle defined by vertexes $(\mathbf{m}_j, \mathbf{m}_k, \mathbf{m}_h)$, while A_1, A_2 and A_3 are the areas of triangles defined by $(\mathbf{m}_j, \mathbf{m}_k, \mathbf{x}_{r,c})$, $(\mathbf{m}_k, \mathbf{x}_{r,c}, \mathbf{m}_h)$ and $(\mathbf{m}_h, \mathbf{m}_j, \mathbf{x}_{r,c})$ respectively (see Figure 4). Finally, each $p_{r,c}$ is converted into the color of the pixel $k_{r,c}$ of a *grayscale image* with the same size of the grid:

$$k_{r,c} = \left\lfloor \frac{255}{\max(P)} p_{r,c} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ is the floor function. The conversion is performed by considering the maximum pressure value measured by the robotic skin for the current contact image. Examples of different contact images generated by a human hand obtained using the described procedure are presented in Figure 2.

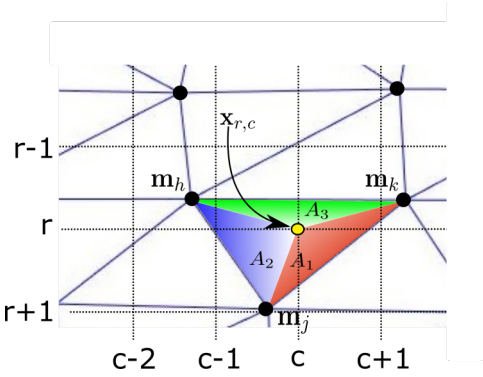


Fig. 4. A regular grid is superimposed on M for contact image formation.

III. Hand Classification

A. Bag of Visual Words

Bag of Visual Word (BoVW) framework [26] is an extension of the Bag of Word (BoW) model, used in text document analysis applications. BoW consists in a histogram representing the word occurrences in a text document. In a similar way, BoVW is defined as a histogram representing *visual word* [26] occurrences in an image. Image classifiers can be trained on these BoVW histograms in order to discriminate different classes of objects. In our application, *hand* and *non-hand* contact images are used to train a BoVW-based machine learning algorithm. The training procedure can be summarized as follows (see Figure 5).

Firstly, *keypoint descriptors* (or features) [27] are extracted from the contact images using robust feature extraction algorithms such as SIFT [27] and SURF [28]. In this work, the latter has been selected because is faster in computing keypoint descriptors[28].

Secondly, keypoint descriptors are clustered, using K-means algorithm, in *visual words*, obtaining a dictionary or a *visual vocabulary*. As a result, this operation groups descriptors with similar characteristics. The performance of the classifier depends on K, i.e. the dictionary size. Indeed, if K is too small, a visual word is too "generic" and not representative of a particular object class. Conversely, a too high value of K could cause overfitting. The proper size of the vocabulary is investigated in the next Section.

Finally, for each image, a histogram of visual words of length K is generated. The set of histograms is used for training a linear Support Vector Machine (SVM) binary classifier on the two classes, i.e., hand and non-hand contacts.

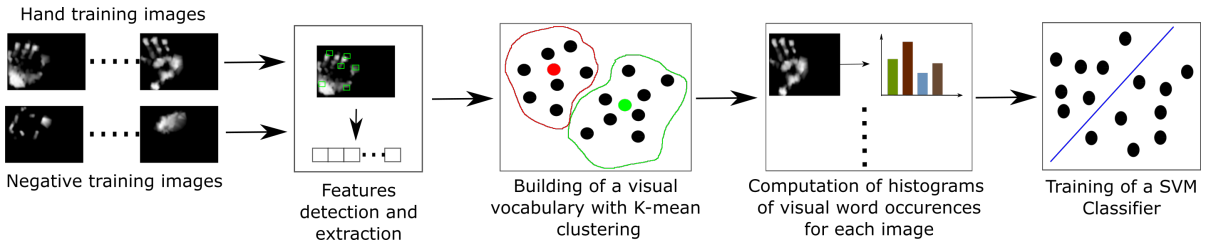


Fig. 5. Description of BoVW training procedure.

B. Convolutional Neural Networks

CNNs are a special kind of neural networks widely used in image classification [29]. The CNN structure has been defined by evaluating the accuracy of different networks on the dataset described in Section IV. The number of layers has been increased until the performance on the training data has been found to be satisfactory, as indicated in [29]. As a result, the CNN structure has been defined as in Figure 6. The first convolutional layer is composed by 50 kernels of 5×5 size. The output of a convolutional layer is elaborated by a Rectified Linear Unit (ReLU) layer that performs a threshold operation on the input. After ReLu, a downsampling with a 2×2 maxpool filter with stride 2 is performed. This structure, composed of Convolutional + ReLU + MaxPooling layers, is repeated three times. The last part of the network consists in a fully-connected layer with a 2-way softmax output unit which computes a probability distribution over the 2 class labels. Our CNN has been trained using stochastic gradient descent algorithm with a momentum of 0.9 and batch size of 128. The learning rate is decreased every 10 training epochs by a 0.0005 factor. Preliminary experiments have shown that the hyperparameters that heavily affect the performance of CNNs in our application are the initial learning rate and the number of training epochs. For this reason, they have been tuned as described in the following Section.

IV. Experimental Results

A. Experimental platform and dataset formation

The dataset required for the experimental evaluation has been produced using the platform described in Section II-A where 768 taxels cover the upper part of the forearm of a Baxter robot (see Figure 3). As discussed before, the contact shape depends on several factors, such as human physical characteristics, manipulation tasks and contact surface shape. In order to generate a dataset that captures this variability, it has been asked to different persons to interact with the robot in both structured and unstructured ways. In particular, it has been asked to a human subject to perform the following actions on the sensorized area of the robot body:

- Grasp the forearm
- Grasp and torque clockwise the forearm
- Grasp and torque counter-clockwise the forearm
- Grasp and push to the left the forearm
- Grasp and push to the right the forearm
- Grasp and push away the forearm
- Pull the forearm toward himself

Each action has been performed in two different predetermined positions. In the former, the subject stands in front of the robot, in the latter, he stands on the robot side (see Figure 7). During the whole experiment, the robot is commanded to maintain its pose. After that, it has been asked to interact freely with the robot in five different ways. During the experiment, the robotic skin feedback has been recorded and a snapshot of the tactile measurements in a steady state condition has been extracted and later used for generating the contact images as explained in Section II. This experiment has been repeated on 43 different human subjects[1] of different gender (60% Male, 40% Female), handedness (81% Right, 19% Left), hand length (17-22 cm), age (22-59 years) and weight (50-100 Kg). The dataset has been completed with about one thousand negative samples i.e. contacts with generic objects (e.g., spheres, cylinders, etc.) or other human body parts (e.g., elbow, forearm and shoulder), resulting in a set of more than 1800 contact images. Each contact image has been resized to 66x100 pixels in order to reduce the training time of machine learning algorithms. The complete dataset is publicly available and can be found in [30].

B. Results

The dataset, formed by *hand* and *non-hand* contacts, has been divided in the training set, 70%, and the test set 30%. The training set has been used for hyperparameter tuning by using a 5-fold cross-validation for each combination of hyperparameter values. The performance in each cross-validation has been evaluated by averaging the results obtained in each validation fold. The parameters related to the best accuracy have been selected.

For BoVW approach, the SURF descriptors length has been set to 64, i.e., a good tradeoff between accuracy and speed [28], while the tuning procedure involved the dictionary size only. Figure 8 shows the result of cross-validation for BoVW. The Figure reports the mean accuracy obtained by classifying the validation folds for increasing values of K. It can be seen that the highest mean accuracy on the validation folds is 97.7% and it has been obtained with a dictionary size of $K = 130$.

Differently, the CNN has been tuned on the number of training epochs and the initial learning rate of the stochastic gradient descent method. The 5-fold cross-validation has been applied with increasing values of initial learning rate

[1] Each subject signed an informed consent form.

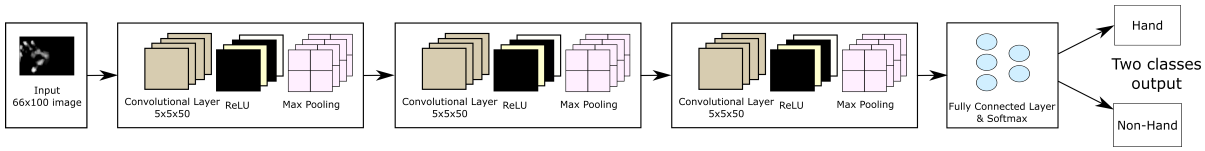


Fig. 6. The CNN structure.

Fig. 7. The two different positions assumed by a human during the experiments: in front of the robot (left) and on its side (right).

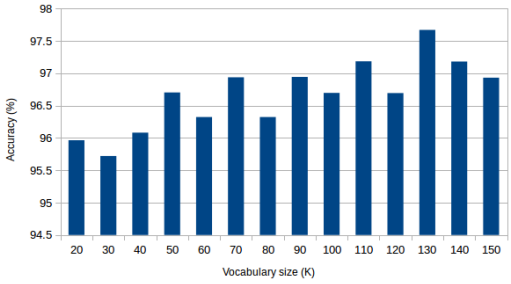


Fig. 8. Accuracy obtained from 5-fold cross-validation procedure for increasing values of the dictionary size K. The highest accuracy is 97.7% and it is obtained for $K = 130$.

and training epochs (see Table I). The results of cross-validation for each combination of the hyperparameters are reported in Table I where it is shown that an initial learning rate of 0.001 and a number of training epochs of 50 provide the best mean accuracy of 98.8% on the validation folds.

After selecting the hyperparameters, both machine learning algorithms have been trained on the training set and the performances have been evaluated on the test set. Table II and Table III report respectively the confusion matrices of the classification performed with BoVW and CNN. BoVW provides a mean accuracy of 97.59% (i.e. mean of the diagonal of the confusion matrix) that is slightly lower than the CNN. However, both methods predict the correct label with a very high accuracy.

Another experiment has been performed in order to test the repeatability of the results shown in Tables II and III. Test and training sets have been randomly generated 10 times and the previously described procedure has been repeated on both methods. The comparison between them can be found in Table IV where statistics on the mean of the diagonal of the two confusion matrices on the 10 trials are reported. Again, CNN provides slightly better prediction accuracy than BoVW. However, both cases present a mean accuracy higher than 98% and the difference in terms of results are minimal. Finally, it should be noted that even if the CNN approach presents a higher classification accuracy, BoVW model structure is simpler than CNN and it has fewer hyperparameters to tune.

V. Conclusions

This paper deals with the problem of using the tactile feedback generated by a robotic skin for discriminating a human hand touch from a generic contact. The proposed solution transforms measurements of pressure sensors into a convenient 2D representation of the contact shape, i.e., a contact image. The aim is to exploit image classification algorithms for the recognition of the human hand touch. Two machine learning techniques have been evaluated and compared, i.e., BoVW and CNNs. A dataset for training and evaluating the two approaches has been built by requesting to 43 different persons to interact with the robot in both structured and unstructured ways. Contacts generated by other body parts or by generic objects have been added obtaining a dataset of more than 1800 contact images. The experimental results show that both algorithms are suited for the task obtaining an accuracy of more than 96% in the worst case. Finally, the algorithm will be integrated into our real-time data acquisition and processing framework for large-scale robot skin [31].

References

[1] J. Krger, T. Lien, and A. Verl, "Cooperation of human and machines in assembly lines," {*CIRP*} *Annals - Manufacturing Technology*, vol. 58, no. 2, pp. 628 – 646, 2009.

[2] H. Kimura, T. Horiuchi, and K. Ikeuchi, "Task-model based human robot cooperation using vision," in *Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289)*, vol. 2, 1999, pp. 701–706 vol.2.

[3] D. M. Ebert and D. D. Henrich, "Safe human-robot-cooperation: image-based collision detection for industrial robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, 2002, pp. 1826–1831 vol.2.

[4] S. Haddadin, A. Albu-Schaffer, A. D. Luca, and G. Hirzinger, "Collision detection and reaction: A contribution to safe physical human-robot interaction," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2008, pp. 3356–3363.

[5] K. Yokoyama, H. Handa, T. Isozumi, Y. Fukase, K. Kaneko, F. Kanehiro, Y. Kawai, F. Tomita, and H. Hirukawa, "Cooperative works by a human and a humanoid robot," in *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, vol. 3, Sept 2003, pp. 2985–2991 vol.3.

[6] G. Cannata, M. Maggiali, G. Metta, and G. Sandini, "An embedded artificial skin for humanoid robots," in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Aug 2008, pp. 434–438.

TABLE I

ACCURACY OBTAINED BY APPLYING 5-FOLD CROSS-VALIDATION WITH DIFFERENT VALUES OF TRAINING EPOCHS AND INITIAL LEARNING RATE FOR CNN.

		Training Epochs														
		10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**110**	**120**	**130**	**140**	**150**
Initial Learning Rate	**1e-5**	61.6%	56.2%	54.7%	55.54%	60.2%	58.1%	63.3%	63.9%	65.6%	71.3%	77.7%	80.3%	83.9%	82.2%	88.4%
	5e-5	53.9%	64.6%	80.8%	89%	95.7%	96.6%	96.8%	96.5%	97.4%	97.3%	97.6%	97.5%	97.7%	97.7%	97.3%
	1e-4	59.5%	84.5%	93.1%	96.8%	97.2%	97.5%	98.1%	98.4%	97.9%	97.5%	97.9%	98.3%	98.1%	98.5%	98%
	5e-4	96.1%	98.1%	98.4%	98.6%	98.4%	97.8%	98.2%	97.9%	97.7%	98.1%	98%	98%	98.1%	98.2%	98.2%
	1e-3	97.3%	98.3%	98.5%	98.2%	**98.8%**	98.2%	98.1%	98.4%	98.4%	98.2%	97.7%	98%	98.4%	98.2%	98.2%
	5e-3	85.5%	96.9%	79.5%	88%	88.3%	79%	88.7%	88.6%	96.1%	88.5%	96.9%	79.2%	97.7%	88.6%	78.6%
	1e-2	65.6%	50.6%	58.3%	49.4%	49.8%	53.2%	57.6%	49.5%	49.4%	52.9%	49.1%	52%	51.7	51.4%	50.7%

TABLE II

CONFUSION MATRIX OF BOVW CLASSIFIER APPLIED ON THE TEST SET. THE MEAN ACCURACY IS 97.59% WITH DICTIONARY SIZE OF 130.

	Hand	**Negative**
Hand	96.92%	3.08%
Negative	1.74%	98.26%

TABLE III

CONFUSION MATRIX OF CNN CLASSIFIER APPLIED ON THE TEST SET. THE MEAN ACCURACY IS 98.68% WITH AN INITIAL LEARNING RATE OF 0.001 AND A NUMBER OF TRAINING EPOCHS OF 50.

	Hand	**Negative**
Hand	98.24%	1.76%
Negative	0.88%	99.12%

TABLE IV

COMPARISON BETWEEN BOVW AND CNN MODELS

	Mean	**Min**	**Max**
BoVW	98.15%	96.92%	98.9%
CNN	98.33%	97.14%	99.34%

[7] P. Mittendorfer and G. Cheng, "Humanoid multimodal tactile-sensing modules," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 401–410, June 2011.

[8] T. Sato, Y. Nishida, J. Ichikawa, Y. Hatamura, and H. Mizoguchi, "Active understanding of human intention by a robot through monitoring of human behavior," in *Intelligent Robots and Systems' 94.'Advanced Robotic Systems and the Real World', IROS'94. Proceedings of the IEEE/RSJ/GI International Conference on*, vol. 1. IEEE, 1994, pp. 405–414.

[9] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot programming by demonstration," in *Springer handbook of robotics*. Springer, 2008, pp. 1371–1394.

[10] H. Liu, J. Greco, X. Song, J. Bimbo, L. Seneviratne, and K. Althoefer, "Tactile image based contact shape recognition using neural network," in *2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, Sept 2012, pp. 138–143.

[11] 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, Sept 2015.

[12] 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, June 2011.

[13] S. Xu, T. Fang, D. Li, and S. Wang, "Object classification of aerial images with bag-of-visual words," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 366–370, April 2010.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[15] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592–3607, Nov 2011.

[16] H. I. Lin, M. H. Hsu, and W. K. Chen, "Human hand gesture recognition using a convolution neural network," in *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, Aug 2014, pp. 1038–1043.

[17] J. Nagi, F. Ducatelle, G. A. D. Caro, D. Cirean, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Nov 2011, pp. 342–347.

[18] H.-J. Kim, J. S. Lee, and J. H. Park, "Dynamic hand gesture recognition using a cnn model with 3d receptive fields," in *2008 International Conference on Neural Networks and Signal Processing*, June 2008, pp. 14–19.

[19] S. S. Farfade, M. J. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. ACM, 2015, pp. 643–650.

[20] "Baxter robot," Available: http://www.rethinkrobotics.com/.

[21] G. Cannata, S. Denei, and F. Mastrogiovanni, "Towards automated self-calibration of robot skin," in *2010 IEEE International Conference on Robotics and Automation*, May 2010, pp. 4849–4854.

[22] A. D. Prete, S. Denei, L. Natale, F. Mastrogiovanni, F. Nori, G. Cannata, and G. Metta, "Skin spatial calibration using force/torque measurements," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2011, pp. 3694–3700.

[23] C. Toth, J. O'Rourke, and J. Goodman, *Handbook of Discrete and Computational Geometry, Second Edition*, ser. Discrete Mathematics and Its Applications. CRC Press, 2004.

[24] G. Cannata, S. Denei, and F. Mastrogiovanni, "Tactile sensing: Steps to artificial somatosensory maps," in *19th International Symposium in Robot and Human Interactive Communication*, Sept 2010, pp. 576–581.

[25] M. Desbrun, M. Meyer, and P. Alliez, "Intrinsic parameterizations of surface meshes," *Computer Graphics Forum*, vol. 21, no. 3, pp. 209–218, 2002.

[26] H. Kato and T. Harada, "Image reconstruction from bag-of-visual-words," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[27] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[28] H. Bay, T. Tuytelaars, and L. Van Gool, *SURF: Speeded Up Robust Features*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417.

[29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[30] "Dataset of contact images," Available: https://github.com/maclab/Contact-Images-Dataset-ROMAN2017.git".

[31] S. Youssefi, S. Denei, F. Mastrogiovanni, and G. Cannata, "A real-time data acquisition and processing framework for large-scale robot skin," *Robotics and Autonomous Systems*, vol. 68, pp. 86–103, 2015.