Tactile Sensing and Machine Learning for Human and Object Recognition in Disaster Scenarios

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

System Engineering and Automation Department, University of Málaga, 29071 Málaga, Spain
{jmgandarias,jesus.gomez,ajgarcia}@uma.es,
https://www.uma.es/robotics-and-mechatronics/

Abstract. This paper presents the application of machine learning to tactile sensing for rescue robotics. Disaster situations often exhibit low-visibility scenarios where haptic feedback provides a valuable information for the search of potential victims. To extract haptic information from the environment, a tactile sensor attached to a lightweight robotic arm is used. Then, methods based on the SURF descriptor, support vector machines (SVM), Deep Convolutional Neural Networks (DCNN) and transfer learning are implemented to classify the data. Besides, experiments have been carried out, to compare those procedures, using different contact elements, such as human parts and objects that could be found in catastrophe scenarios. The best achieved accuracy of 92.22%, results from the application of the transfer learning procedure using a pre-trained DCNN and fine-tuning the classification layer of the network.

Keywords: Tactile Sensors, Rescue Robotics, Machine Learning, Deep Learning, Transfer Learning, Object Recognition

1 Introduction

Teleoperation is still a critical element in rescue robotics due to the complexity of the operations in an unstructured environment [15]. Previous experiences in real situations have evidenced the problematic of using systems provided with visual perception only. In low-light scenarios, or in presence of dust or smoke, systems with haptic feedback contribute with additional information that can compensate the lack of visual information [24].

A key task in rescue robotics operations with large number of victims, is to locate and evaluate the urgency degrees of the victims, in function of the priority of the treatment *(triage)*. This task raise technological problems such as Human-Robot Interaction *(HRI)*, which is considered one of the biggest challenges in this field [20]. A first approach to achieve a solution would be the identification of the victims and the different parts of the body, prior to the measurement of the vital signs.

Identifying potential victims in disaster scenarios is a priority for search and rescue teams, but it still an research problem. In [26], typical problems of locating and tracking humans with video segmentations are presented. Also, a proposal to enhance the visual tracking by fusing both color and thermal-infrared spectra is described. Another approach [6], is based on the abrupt changes suffered by the UWB signals characteristics when passing through a human body, to simultaneously detect human beings and locate the rescue robot using UWB signals.

Haptic feedback can benefit both robotic guidance and victims manipulation. In robotic guidance, haptics systems enhance the control of the teleoperator, who can drive the vehicle with force feedback based on potential fields. This perception reduces the visual-information dependence, driving the operator to the target [1], avoiding collisions, or combining both functions [2]. Other application of haptic technology to rescue robots are based on the use of a guide vehicle in low-visibility situations [24]. The University of Málaga has contributed previously to the application of tactile sensors to rescue-robots [28], where a tactile sensor was developed to provide a pressure map of the external applied forces, and this sensor was attached to an end-hydraulic-effector of a rescue robot [7].

Different tactile sensing implementations can't be easily compared, due to the differences of the hardware [21]. One approach consist on using tactile data to recognize objects by their shape [14]. On the other hand, interpreting tactile data as time series was also investigated [19], [13]. The majority of these works are based on artificial intelligence algorithms to classify the data. One proposition consists on employing computer vision algorithms and machine learning techniques [17], [16], whilst other employs neural networks [21], [9] and deep learning. For instance, [25] presents the use of Deep Learning with Dropout to reduce overfitting, and the benefits of including both kinesthetic and tactile information to object shape recognition. [18] also present the benefits of using both kinesthetic and tactile information to recognize objects. Other approach of using Deep Learning techniques and artificial tactile sensing consists on determining the contact material [3].

This work proposes the use of tactile sensors in emergency situations, where searching and rescuing potential victims are a priority. A system composed by a lightweight robotic manipulator and a high-resolution tactile sensor is employed to recognize objects. Moreover, machine learning methods for pressure image recognition are presented. These methods are based on two steps: extracting features from the pressure-tactile images, and classifying those images into predefined labels. Due to the shortage of features of the tactile images, using deep convolutional neural networks (DCNN) [11] has been considered. First, a method based on the *Speeded-Up Robust Features* (SURF) [4] and a *support vector machine* (SVM) [5] (SURF-SVM) is implemented as a checkpoint for the posterior comparison with the DCNN-based methods. One proposal consists on using the features-extraction layers of a pre-trained DCNN as a feature extractor and a SVM to get a classifier (DCNN-SVM method). Besides, a transfer learning procedure with a pre-trained DCNN is also tested. Finally, an experiment with 9 classes is carried out to evaluate and compare the methods in terms of accuracy rate. These experiments include objects that can easily encountered in catastrophes situations and parts of human arms.

The rest of the document is structured as follows. In the section 2, methods characteristics and schemes are described. Thereafter, section 3 shows the experimental procedure, the methodology implementation considerations, and the data and results obtained. Finally, the conclusions and future work are detailed in section 5.

2 Methods

2.1 SURF-SVM

SURF-SVM method applies the SURF descriptor as a feature extractor from pressure images, and a SVM as a classifier. Fig. 1 presents the procedure of the SURF-SVM method implementation. SURF brings out the pressure-tactile images descriptors, which are then clustering in a bag of words framework (BoW) [22] by a k-means algorithm [12], forming a dictionary. Finally, a supervised SVM is trained using the dictionary generated and the known images labels.

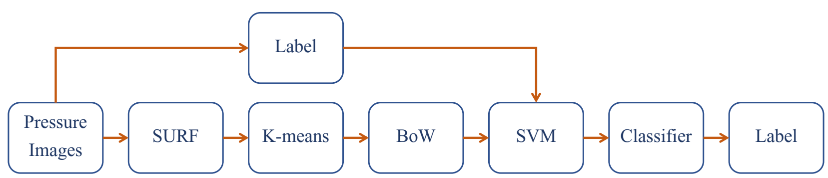

Fig. 1: SURF-SVM scheme.

2.2 DCNN-SVM

This method is similar to the previous one. Nevertheless, the DCNN-SVM replaces the SURF descriptor by the feature extraction part of a DCNN in order to reinforce the lack of information presented in pressure images. This DCNN has been previously pre-trained to classify visual images taken with a normal camera. Hence, we can divide the DCNN architecture in two parts: features extraction, and classification. Fig. 2 shows the DCNN-based algorithm, which uses the activations of the last feature extraction layer of the DCNN and a subsequent SVM, trained in base on these activations and the known labels. The activations used for training the SVM represent the descriptive information of the images.

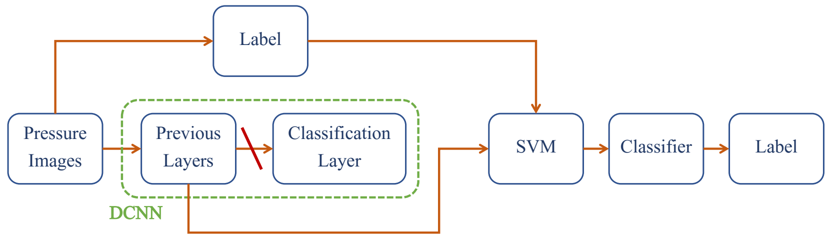

Fig. 2: DCNN-SVM scheme.

2.3 Transfer learning

Transfer learning takes uses a Neural Network trained in an specific domain, with big amount of data, using this network for another purpose in which that volume of data is not available [23]. In our case, we have employed a DCNN that has been previously trained for image classification. As mentioned before, the last section of the DCNN incorporates the classification procedure. The classification section has been trained to classify pressure images. The scheme of this method is showed in Fig 3.

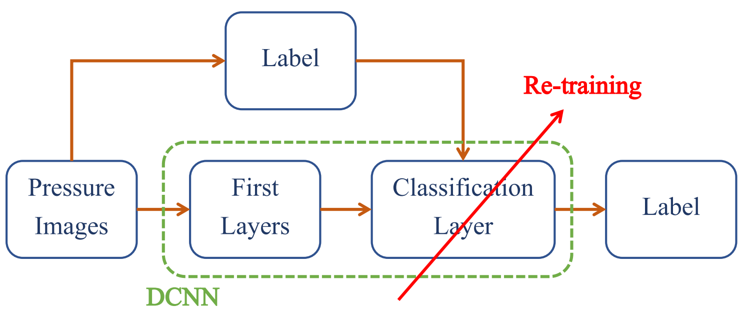

Fig. 3: Transfer learning scheme.

3 Experimental setup

To carry out the experiments, the tactile sensor, model 60077 from *Tekscan*, has been attached to the lightweight robotic manipulator *AUBO OUR-i5*. Fig. 4 shows the experimental setup. The high-resolution resistive tactile-array has a total of 1400 pressure sensels. Each sensel size is 53.3mm x 95.3mm and the sensor presents a density of 27.6 sensels/cm^2 distributed in a matrix composed by 28 rows and 50 columns. Further, a 2mm silicone rubber cover has been added to receive external forces and protect the sensor film.

Fig. 4: Tactile sensor attached to the *AUBO OURi5* robotic manipulator.

In order to evaluate the methods, a collection of 450 pressure images from different elements has been obtained. The data set consists of the raw pressure data given by the tactile sensor in the form of pressure images. These images are picked up from human body parts and inert objects, forming a set of 9 classes: Fingers, Forearm, Hand, Cable Pipe, Rocks, Rubble, Timber Wood, Branch Wood and Piece of Wood.

Fig. 5 shows an example of the body parts and their corresponding pressure maps. On the other hand, examples of each inert-object class and their respective pressure maps are shown in Fig. 6. Those parts of the body are chosen to keep the comfort of a suppose victim, and the ease of access and the manipulability of the robot. Likewise, the inert objects selected are elements that could be found in a disaster scenario.

The learning procedure, also called training, of the proposed methods employs a subset of 180 images of the collected data. After training the methods, a test phase is performed to validate the results, which employs a subset of 270 images.

For the DCNN-SVM method, a large variety of neural networks can be used. To show the differences between them, three DCNNs have been implemented for the application of this method: AlexNet [10], VGG-16 and VGG-19 [27], named as ANET-SVM, VGG16-SVM and VGG19-SVM, respectively. The Neural Networks have been obtained from the *Caffe* repository [8].

Additionally, for the transfer learning application, the DCNN AlexNet has been implemented. That configuration has been named as TL-ANET. In this

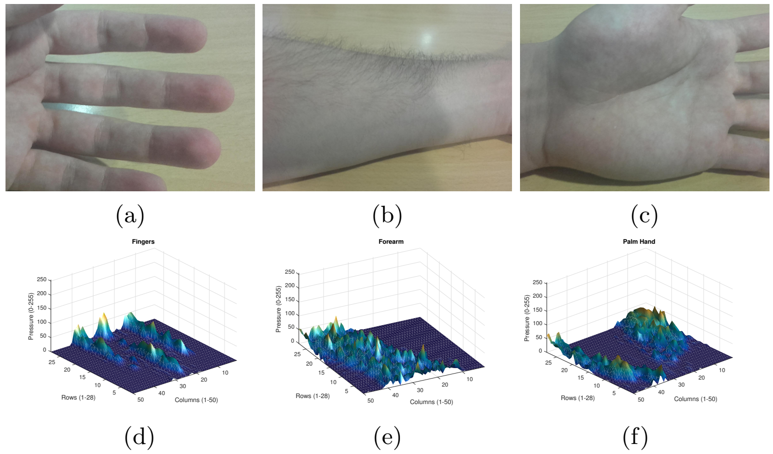

Fig. 5: Parts of the body used for the experiments, labeled as Hand (a), Forearm (b) and Fingers (c)) and their corresponding pressure map ((d), (e), (f).

method, only the classification layer has been re-trained using a single GTX 1050Ti GPU with 4GB of memory.

4 Results

Fig. 7 shows the resulting confusion matrices of the methods, whilst Table 1 presents the summary of the results. With the application of the SURF-SVM method, a recognition rate of 70.74% has been achieved. An improvement of the accuracy, produced by the application of Deep Learning techniques, in extracting features from pressure images has been detected in contrast with the SURF-based method. In particular, DCNN-SVM methods (ANET-SVM, VGG16-SVM and VGG19-SVM) shows an improvement between 13.7% and 17.04%. Furthermore, a fine-tune of the classification layer of the DCNN (TL-ANET) implies an improvement of 21.48% against the use of an SVM (SURF-SVM) to classify the data.

5 Conclusions

An application to the detection of victims in disaster scenarios, based on tactile sensing, has been presented. A high-resolution tactile sensor provides raw pressure data distribution. Due to the lack of information brought by the pressure images, deep learning techniques have been considered and compared with respect to a non-deep-learning algorithm (SURF-SVM), which implements the SURF descriptor as a feature extractor and a SVM to get a classifier. Further, three pre-trained DCNN have been employed as feature extractors, exchanging the last classification layer by a SVM (DCNN-SVM). Additionally, a transfer learning procedure to re-train the last layer of the AlexNet DCNN (TL-ANET) in a single GPU is also developed. Besides, the lightweight robotic manipulator *AUBO OUR-i5* provided with a tactile sensor has been used to capture the pressure images for carrying out a 9-classes experiment. The 9-classes are distributed into 3-classes for human body parts (forearm, hand and fingers) and 6-classes for inert objects that could be found in a catastrophe scenario (cable pipe, rubble, rocks, timber, branch wood and a small piece of wood). Results of the experiments reveal an accuracy improvement between 13.7% and 17.04% using the DCNN-SVM method with respect to the 70.74% accuracy of the SURF-SVM, and an improvement of 21.48% achieved with the (TL-ANET), showing that a recognition rate of 92.22% for distinguishing human from objects can be achieved using transfer learning. In future work, time-series data combined with active touch and palpation strategies will be considered. Also, other information such as the relation between the applied forces and the displacement of the sensor with respect to the contact with an element will be used.

Table 1: Summary of results

Method	**Accuracy [%]**	**Improvement [%]**
SURF-SVM	70.74	-
ANET-SVM	84.44	13.7
VGG16-SVM	86.11	15.37
VGG19-SVM	87.78	17.04
TL-ANET	92.22	21.48

Acknowledgment

This work was partially supported by the Spanish project DPI2015-65186-R and the European Commission under grant agreement BES-2016-078237.

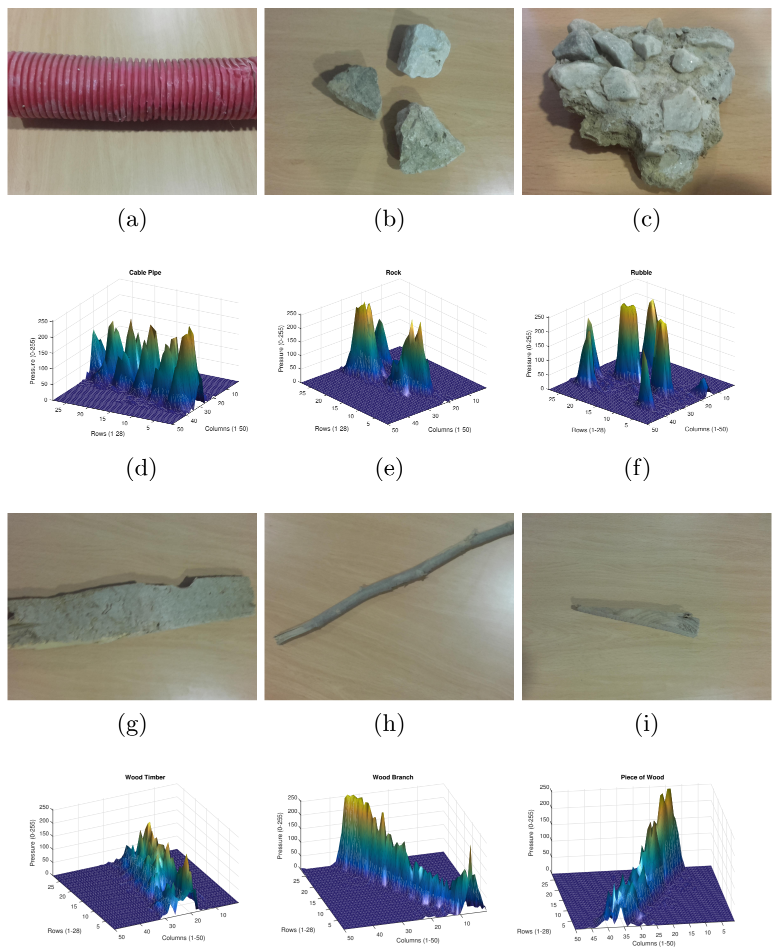

Fig. 6: Inert objects used for the experiments, labeled as Cable pipe (a), Rocks (b), Rubble (c), Timber Wood (g), Branch Wood (h) and Piece of Wood (i), and their corresponding pressure maps (d), (e), (f), (j), (k), (l).

	Arm	Fingers	Hand	Pipe	Wood	Timber	Branch	Rock	Rubble
Arm	0.47	0.00	0.00	0.00	0.17	0.00	0.13	0.13	0.10
Fingers	0.00	0.87	0.00	0.00	0.00	0.00	0.07	0.07	0.00
Hand	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.00	0.00
Pipe	0.00	0.03	0.03	0.90	0.00	0.00	0.03	0.00	0.00
Wood	0.17	0.00	0.00	0.00	0.40	0.07	0.07	0.17	0.13
Timber	0.00	0.00	0.00	0.00	0.00	0.97	0.03	0.00	0.00
Branch	0.00	0.00	0.03	0.00	0.00	0.20	0.77	0.00	0.00
Rock	0.10	0.00	0.07	0.03	0.07	0.00	0.00	0.70	0.03
Rubble	0.13	0.00	0.00	0.00	0.30	0.03	0.03	0.04	0.43

(a)

	Arm	Fingers	Hand	Pipe	Wood	Timber	Branch	Rock	Rubble
Arm	0.95	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
Fingers	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hand	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Pipe	0.00	0.00	0.05	0.95	0.00	0.00	0.00	0.00	0.00
Wood	0.00	0.00	0.00	0.00	0.75	0.00	0.00	0.10	0.15
Timber	0.00	0.00	0.00	0.00	0.05	0.80	0.15	0.00	0.00
Branch	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Rock	0.00	0.00	0.30	0.00	0.05	0.00	0.00	0.65	0.00
Rubble	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.50

(b)

	Arm	Fingers	Hand	Pipe	Wood	Timber	Branch	Rock	Rubble
Arm	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fingers	0.00	0.95	0.00	0.00	0.00	0.00	0.00	0.05	0.00
Hand	0.00	0.00	0.95	0.05	0.00	0.00	0.00	0.00	0.00
Pipe	0.00	0.00	0.05	0.95	0.00	0.00	0.00	0.00	0.00
Wood	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.20	0.55
Timber	0.00	0.00	0.00	0.00	0.00	0.95	0.05	0.00	0.00
Branch	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Rock	0.00	0.05	0.05	0.00	0.10	0.00	0.00	0.80	0.00
Rubble	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.90

(c)

	Arm	Fingers	Hand	Pipe	Wood	Timber	Branch	Rock	Rubble
Arm	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fingers	0.00	0.95	0.00	0.00	0.00	0.00	0.00	0.05	0.00
Hand	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Pipe	0.00	0.05	0.05	0.90	0.00	0.00	0.00	0.00	0.00
Wood	0.00	0.00	0.00	0.00	0.35	0.10	0.00	0.20	0.35
Timber	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Branch	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Rock	0.00	0.15	0.05	0.00	0.05	0.00	0.00	0.75	0.00
Rubble	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.95

(d)

	Arm	Fingers	Hand	Pipe	Wood	Timber	Branch	Rock	Rubble
Arm	0.95	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Fingers	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hand	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Pipe	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Wood	0.00	0.00	0.00	0.00	0.75	0.05	0.00	0.05	0.15
Timber	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Branch	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Rock	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.90	0.05
Rubble	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.70

(e)

Fig. 7: Confusion matrix resulting of applying the methods: SURF-SVM (a), ANET-SVM (b), VGG16-SVM (c), VGG19-SVM (d), TL-ANET (e).

References

[1] Abdelrahman, W., Wei, L., Mullins, J., Nahavandi, S.: Wireless haptic rendering for mobile platforms. In: Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, pp. 2213–2217. IEEE (2012). DOI 10.1109/ICSMC.2012.6378069

[2] Ang, Q.Z., Horan, B., Nahavandi, S.: Multipoint haptic mediator interface for robotic teleoperation. IEEE Systems Journal **9**(1), 86–97 (2015). DOI 10.1109/JSYST.2013.2283955

[3] Baishya, S.S., Bauml, B.: Robust material classification with a tactile skin using deep learning. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8–15. IEEE (2016). DOI 10.1109/IROS.2016.7758088

[4] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding **110**(3), 346–359 (2008). DOI 10.1016/j.cviu.2007.09.014

[5] Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3), 273–297 (1995). DOI 10.1007/BF00994018

[6] Espes, D., Pistea, A.M., Canaff, C., Iordache, I., Parc, P.L., Radoi, E.: New Method for Localization and Human Being Detection using UWB Technology: Helpful Solution for Rescue Robots. arXiv preprint arXiv:1312.4162 (2013)

[7] García-Cerezo, A., Mandow, A., Martínez, J.L., Gómez-de Gabriel, J., Morales, J., Cruz, A., Reina, A., Serón, J.: Development of ALACRANE: A mobile robotic assistance for exploration and rescue missions. In: SSRR2007 - IEEE International Workshop on Safety, Security and Rescue Robotics Proceedings, pp. 1–6. IEEE (2007). DOI 10.1109/SSRR.2007.4381269

[8] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe. Proceedings of the ACM International Conference on Multimedia - MM '14 **1436**, 675–678 (2014). DOI 10.1145/2647868.2654889

[9] 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. In: The 2012 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2012). DOI 10.1109/IJCNN.2012.6252593

[10] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105 (2012)

[11] LeCun, Y., Bengio, Y., et al.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks **3361**(10), 1995 (1995)

[12] Likas, A., Vlassis, N., Verbeek, J.J.: The global k-means clustering algorithm. Pattern recognition **36**(2), 451–461 (2003)

[13] Liu, H., Guo, D., Sun, F.: Object Recognition Using Tactile Measurements: Kernel Sparse Coding Methods. IEEE Transactions on Instrumentation and Measurement **65**(3), 656–665 (2016). DOI 10.1109/TIM.2016.2514779

[14] 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. In: Proceedings - IEEE International Conference on Robotics and Automation, pp. 1410–1415. IEEE (2012). DOI 10.1109/ICRA.2012.6224872

[15] Liu, Y., Nejat, G.: Robotic Urban Search and Rescue: A Survey from the Control Perspective. Journal of Intelligent & Robotic Systems **72**(2), 147–165 (2013). DOI 10.1007/s10846-013-9822-x

[16] Luo, S., Liu, X., Althoefer, K., Liu, H.: Tactile object recognition with semi-supervised learning. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9245, pp. 15–26. Springer, Cham (2015). DOI 10.1007/978-3-319-22876-1_2

[17] Luo, S., Mou, W., Althoefer, K., Liu, H.: Novel Tactile-SIFT Descriptor for Object Shape Recognition. IEEE Sensors Journal **15**(9), 5001–5009 (2015). DOI 10.1109/JSEN.2015.2432127

[18] Luo, S., Mou, W., Althoefer, K., Liu, H.: Iterative Closest Labeled Point for Tactile Object Shape Recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE (2016). DOI 10.1109/IROS.2016.7759485

[19] Madry, M., Bo, L., Kragic, D., Fox, D.: ST-HMP: Unsupervised Spatio-Temporal feature learning for tactile data. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 2262–2269. IEEE (2014). DOI 10.1109/ICRA.2014.6907172

[20] Murphy, R.R., Tadokoro, S., Nardi, D., Jacoff, A., Fiorini, P., Choset, H., Erkmen, A.M.: Search and Rescue Robotics. In: Springer Handbook of Robotics, pp. 1151–1173. Springer (2008). DOI 10.1007/978-3-540-30301-5_51

[21] Navarro, S.E., Gorges, N., Wörn, H., Schill, J., Asfour, T., Dillmann, R.: Haptic object recognition for multi-fingered robot hands. In: Haptics Symposium 2012, HAPTICS 2012 - Proceedings, pp. 497–502. IEEE (2012). DOI 10.1109/HAPTIC.2012.6183837

[22] Nowak, E., Jurie, F., Triggs, B.: Sampling Strategies for Bag-of-Features. Eccv **3954**, 490–503 (2006). DOI 10.1007/11744085_38

[23] Pan, S.J., Yang, Q.: A survey on transfer learning (2010). DOI 10.1109/TKDE.2009.191

[24] Ranasinghe, A., Sornkarn, N., Dasgupta, P., Althoefer, K., Penders, J., Nanayakkara, T.: Salient Feature of Haptic-Based Guidance of People in Low Visibility Environments Using Hard Reins. IEEE Transactions on Cybernetics **46**(2), 568–579 (2016). DOI 10.1109/TCYB.2015.2409772

[25] Schmitz, A., Bansho, Y., Noda, K., Iwata, H., Ogata, T., Sugano, S.: Tactile object recognition using deep learning and dropout. 2014 IEEE-RAS

International Conference on Humanoid Robots pp. 1044–1050 (2014). DOI 10.1109/HUMANOIDS.2014.7041493
[26] Serrano-Cuerda, J., Lopez, M.T., Fernández-Caballero, A.: Robust human detection and tracking in intelligent environments by information fusion of color and infrared video. In: Proceedings - 2011 7th International Conference on Intelligent Environments, IE 2011, pp. 354–357. IEEE (2011). DOI 10.1109/IE.2011.21
[27] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
[28] Vidal-Verdú, F., Barquero, M.J., Castellanos-Ramos, J., Navas-González, R., Sánchez, J.A., Serón, J., García-Cerezo, A.: A large area tactile sensor patch based on commercial force sensors. Sensors **11**(5), 5489–5507 (2011). DOI 10.3390/s110505489