Multi-modal haptic image recognition based on deep learning

Dong Han, Hong Nie and Jinbao Chen
State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing, China
Meng Chen
Aerospace System Engineering Shanghai, Shanghai, China, and
Zhen Deng and Jianwei Zhang
Department of Informatics, Institute of Technical Aspects of Multimodal Systems, University of Hamburg, Hamburg, Germany

Abstract

Purpose – This paper aims to improve the diversity and richness of haptic perception by recognizing multi-modal haptic images.

Design/methodology/approach – First, the multi-modal haptic data collected by BioTac sensors from different objects are pre-processed, and then combined into haptic images. Second, a multi-class and multi-label deep learning model is designed, which can simultaneously learn four haptic features (hardness, thermal conductivity, roughness and texture) from the haptic images, and recognize objects based on these features. The haptic images with different dimensions and modalities are provided for testing the recognition performance of this model.

Findings – The results imply that multi-modal data fusion has a better performance than single-modal data on tactile understanding, and the haptic images with larger dimension are conducive to more accurate haptic measurement.

Practical implications – The proposed method has important potential application in unknown environment perception, dexterous grasping manipulation and other intelligent robotics domains.

Originality/value – This paper proposes a new deep learning model for extracting multiple haptic features and recognizing objects from multi-modal haptic images.

Keywords Deep learning, Data fusion, Haptic perception, Multi-modal haptic images

Paper type Research paper

1. Introduction

Haptic recognition is one of the major issues in robotics manipulation (Chitta *et al.*, 2011; Dargahi and Najarian, 2004), medical diagnostics (Arian *et al.*, 2014), prosthetics (Cu *et al.*, 2016) and haptic display (Tian *et al.*, 2016, 2017). Although there have been lots of studies on vision-based recognition in recent years (Abdulnabi *et al.*, 2015; Chen *et al.*, 2013), it is difficult to infer many properties of objects from vision alone in some special scenarios (Zhang *et al.*, 2017). Therefore, this paper focuses on developing a novel haptic recognition method.

A wide variety of technologies have been presented for haptic recognition recently. Song *et al.* (2014) designed a novel fabric surface texture sensor using polyvinylidene fluoride film. Orii *et al.* (2017) extracted tactile textures from the time-series data of a pressure sensor and a six-axis acceleration sensor using Convolutional Neural Networks (CNNs). Gorges *et al.* (2010) developed a planar tactile sensor matrix for perceiving object shapes. Zhang *et al.* (2017) described a Monte-Carlo-Tree-Search-based algorithm for actively selecting a sequence of end-effector poses to recognize objects. These methods mainly focus on individual sensing modalities rather than the multi-modal combination of sensory capabilities found in human skin.

BioTac, a multi-modal biomimetic sensor developed by SynTouch LLC, can measure force, micro-vibration and thermal flux simultaneously, which provides a better solution to multi-modal haptic perception (Han *et al.*, 2016; Fishel and Loeb, 2012a; Lin *et al.*, 2009). Fishel and Loeb (2012b) proposed a Bayesian exploration method that can adaptively select the optimal exploratory movement to accurately discriminate textures with BioTac. Wettels and Loeb (2011) designed an artificial neutral network to extract multiple haptic features from the BioTac's output. Chu *et al.* (2013,2015) used

The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/0260-2288.htm**

Sensor Review

[DOI 10.1108/SR-08-2017-0160]

The authors would like to thank the anonymous reviewers for their critical and constructive review of the manuscript. This study was supported by Funding of Jiangsu Innovation Program for Graduate Edu- cation (no. KYLX16_0388), the Fundamental Research Funds for the Central Universities, National Natural Science Foundation of China (no. 51675264) and Open Foundation of Shanghai Key Laboratory of Spacecraft Mechanism.

Received 14 August 2017
Revised 15 September 2017
7 December 2017
6 January 2018
Accepted 6 January 2018

Hidden Markov Models to recognize 25 binary haptic adjectives using the BioTac's data from several exploratory procedures (EPs). Based on Chu's research, Gao *et al.* (2016) proposed a multi-modal CNN model to fuse haptic and visual inputs for recognizing haptic adjectives. However, the studies of Chu and Gao are essentially a binary classification task, which only outputs a "yes" or "no" answer and cannot provide richer features.

Therefore, this paper combines the multi-modal haptic signals from BioTac into haptic images, and builds a new multi-class and multi-label haptic image recognition model based on CNN. This recognition model is able to extract four haptic features, which are hardness, thermal conductivity, roughness and texture, and recognize objects (Figure 1). Compared with the methods by Chu *et al.* (2013, 2015) and Gao *et al.* (2016), the main innovation of the proposed model is that it can simultaneously output multiple haptic features and multiple classes for each feature instead of only single binary feature, which is conducive to more diversified and richer haptic perception.

2. Haptic data set

The haptic data set used in this paper is Penn Haptic Adjective Corpus 2 (PHAC-2) (Chu *et al.*, 2015). PHAC-2 contains the haptic data of 60 different household objects, which are classified into eight classes according to their materials: *Foam*, *Organic*, *Fabric*, *Plastic*, *Paper*, *Stone*, *Glass* and *Metal* (Figure 2).

The haptic data of each object were collected by two BioTac sensors that were mounted on the gripper of a Personal Robot 2 (PR2) robot. For each object, the measurement was performed in ten trials, and each trial included four EPs: *Squeeze*, *Hold*, *Slow Slide* and *Fast Slide* (Chu *et al.*, 2015).

BioTac outputs five types of signals (Figure 3): low-frequency fluid pressure (PDC), high-frequency fluid vibration (PAC), core temperature (TDC), change rate of the core temperature (TAC), 19 electrodes (E1,..., E19) spatially distributed on the surface of the rigid core (Fishel, 2012).

The two BioTac sensors are treated as two distinct instances for augmenting the haptic data set. Therefore, the total number of haptic instances in PHAC-2 is now 1,200 (60 objects × 2 BioTacs× 10 trials).

3. Haptic data pre-processing

3.1 Setting labels

Labels are essential for training and testing deep network models. There are two kinds of labels: object category labels and haptic feature labels in this paper.

In PHAC-2, *Glass* and *Stone* only include two objects each, which are too few for deep learning. Therefore, the four objects of these two categories are deleted, which means 56 objects, namely, six categories, are retained, as shown in Table I.

According to the BioTac's output, the haptic features of objects in this paper include hardness, thermal conductivity, roughness and texture, and each feature is described by more than two classes. The hardness of objects is divided into six

Figure 1 Illustration of the proposed algorithm framework

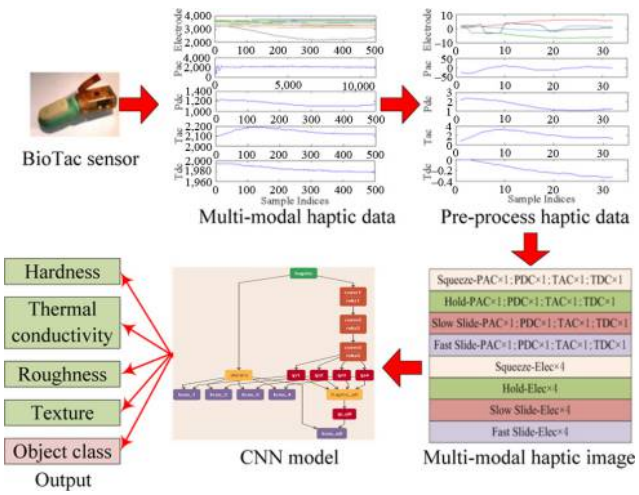


Figure 3 The output signals of BioTac

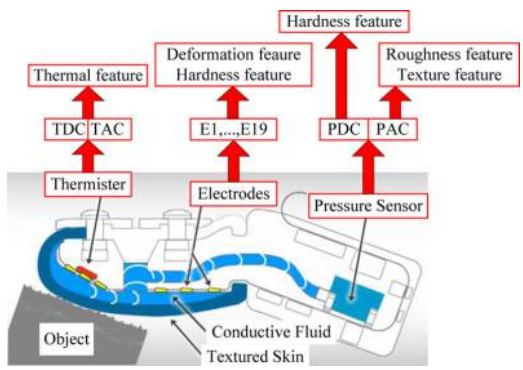


Table I Labels for object categories

Object category	Organic	Plastic	Paper	Fabric	Foam	Metal
Label	0	1	2	3	4	5

Figure 2 The 60 objects in PHAC-2

Foam	Black Foam	Applicator Pad	Charcoal Foam	Bumpy Foam	Gray Foam	Kitchen Sponge	Pool Noodle	Orange Sponge	Pink Foam	Furry Eraser	Blue Sponge	Koozie	Shelf Liner	White Foam	Yellow Foam	Flat Foam
Plastic	Black Acrylic	Bath Cloth	Cutting Board	Caliper Case	Machined Plastic	Plastic Dispenser	Soap Dispenser	Sawed Plastic	Silicone Block	Card Case	Bubble Wrap	Plastic Case	Tarp			
Paper	Colorful Book	Note-book	Envelope	Cosmetics Box	Notepad	RedTooth paste	BlueTooth paste	Tissue Pack	Toilet Paper	Fiber board	Cookie Box	Pen Case				
Fabric	Cloth Sack	Black Eraser	Placemat	Gray Eraser	Yellow Felt	SatinPill owcase	Dish cloth									
Organic	Cork-board	Coco Liner	Plly-wood	Loofah	Layered Cork											
Metal	Aluminum Channel	Aluminum Block	Steel Vase													
Stone	Concrete	Brick														
Glass	Glass Container	Glass Bottle														

classes from the softest to the hardest. Thermal conductivity is divided into six classes from the lowest conductivity to the highest conductivity. Roughness is divided into four classes from the smoothest to the roughest. Texture is divided into three classes from the coarsest to the finest, which are used for describing the fineness of the surface textures of objects. The labels of these four features were set for each object mainly according to the change rate of PDC signal, the change of TAC signal, the power of PAC signal and the spectral center frequency of PAC signal, respectively (Chu *et al.*, 2015; Fishel and Loeb, 2012b). Moreover, the known information of objects also contributed to the objective label setting.

3.2 Normalization

The haptic signals collected by BioTac have different ranges. Therefore, the haptic data should be normalized before being inputted into deep network models, as follows:

$$\boldsymbol{s}'_{ij} = \frac{\boldsymbol{s}_{ij} - \bar{s}_{ij}}{\sigma_{ij}} \tag{1}$$

where $\boldsymbol{s}_{ij}$ is the ith type of haptic signal across all objects in all jth-EPs; i = 1, 2,. . ., 5, which means PAC, PDC, TAC, TDC and E1,. . ., E19, respectively; j = 1, 2, 3, 4, which means *Squeeze*, *Hold*, *Slow Slide* and *Fast Slide*, respectively; $\bar{s}_{ij}$ represents the mean of $\boldsymbol{s}_{ij}$; σ_{ij} is the standard deviation of $\boldsymbol{s}_{ij}$; $\boldsymbol{s}'_{ij}$ is the normalized signal.

3.3 Dimension reduction for electrode signals

As mentioned above, there are 19 electrodes in BioTac. However, experiments found that four principal components capture 95 per cent of the variations of E1,. . ., E19 (Gao *et al.*, 2016). Consequently, the 19 dimensional electrode signals are transformed into four dimensions by the Principal Component Analysis (PCA) algorithm (Abdi and Williams, 2010), which can extract the main change characteristics, reduce useless data and improve the efficiency of training deep network models.

3.4 Data augmentation

The total number of haptic instances is only 1,120 (56 objects × 2 Biotacs × 10 trials). Apparently, it is a small data set for training a deep network model. To augment the data set, each signal is sub-sampled five times at different starting points to obtain five different signals (Figure 4). After the data augmentation, the number of haptic instances is increased to 5,600.

3.5 Making signals with the same length

PAC is sampled at a much higher frequency than other signals for getting high-frequency vibration, leading to its far longer length. Furthermore, the signal lengths in *Squeeze* vary considerably among objects due to different hardness. The haptic signals with different lengths cannot be directly used for haptic images. Therefore, the following method is adopted to make the signals with the same length:

- The minimum signal length $l_{\min}$ is found among the signals gotten from all jth-EPs.
- PAC is divided into three parts (Figure 5), first; then, the *Start* and *End* parts are removed because these two parts are easily disturbed by external noise; finally, the Stable Vibration part is sub-sampled to length of $l_{\min}$. Compared with directly sub-sampling PAC to length of $l_{\min}$, this method can retain the valuable data of PAC as much as possible.
- For other signals, sub-sampling will destroy their gradient features because of different sampling ratios. Therefore, from the start of the signals except PAC, the parts with length of $l_{\min}$ are retained (Figure 6).

According to the requirement of haptic images (Section 3.6), all haptic signals are then compressed into a fixed length by sub-sampling.

An example of pre-processing haptic signals is shown in Figure 7, where the raw signals are from one object in a *Squeeze* procedure. In Figure 7, taking PDC as an example, the raw PDC signal is normalized according to equation (1) first; then, the normalized signal is augmented five times, making the length change from 750 to 150; third, the new PDCs are cut to length of $l_{\min}$= 95; finally, the signals are compressed into length of 32 to construct a haptic image.

3.6 Building multi-modal haptic images

To form the input of CNNs, the processed haptic signals from one trial are concatenated into haptic images with dimension of $C \times T$, where C is the number of signal channels, namely, the image height, and T is the number of time steps, namely, the image length (Figure 8).

The haptic images in Figure 8(*a*) have 32 channels [(PAC + PDC + TAC + TDC + 4 Electrodes) × 4 EPs], which contain all the haptic signals from one trial. To explore the influence of image lengths on haptic image recognition, the three haptic

Figure 4 Illustration of data augmentation

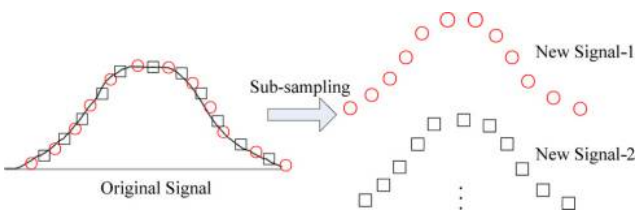


Figure 5 Dividing PAC into three parts: start, stable vibration, end

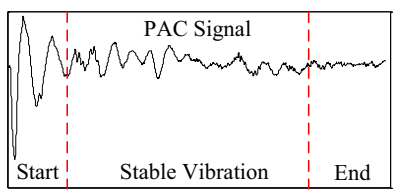


Figure 6 Illustration of cutting the signals to length of $l_{\min}$ in squeeze procedures

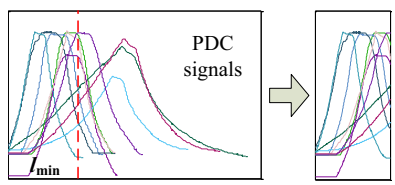


Figure 7 An example of pre-processing haptic signals, the left is the raw signals from one object in a squeeze procedure, the right is the pre-processed signals, PDC is used as an example for illustrating the pre-processing procedures

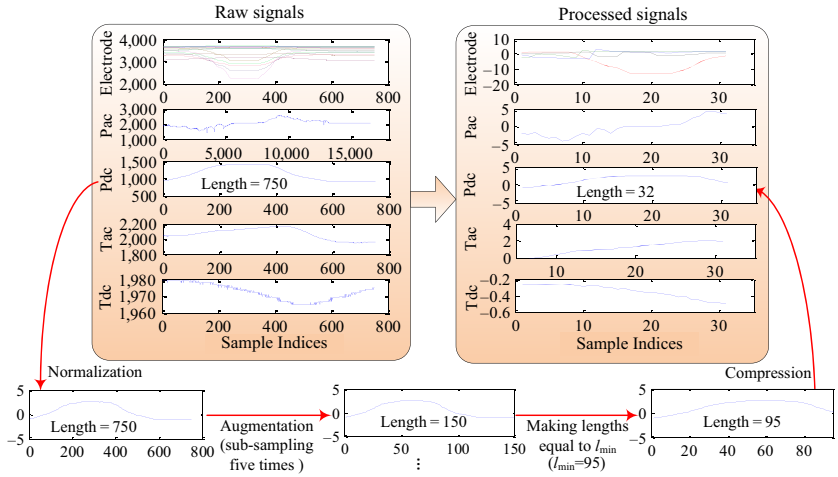
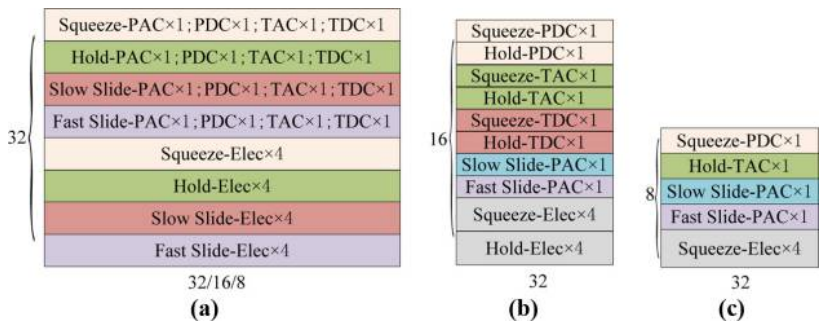


Figure 8 Haptic images with different dimensions, where (a) shows the images with dimensions of 32 × 32, 32 × 16, 32 × 8; (b) and (c) show the images with dimensions of 16 × 32 and 8 × 32, respectively; Elec represents the electrode signals



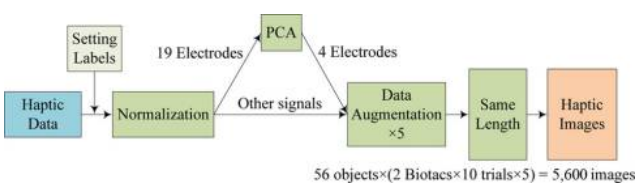
images with 32 channels are set with different lengths: 32 × 32, 32 × 16, 32 × 8. The haptic images in Figure 8(*b*) and 8(*c*) contain 16 channels and 8 channels, respectively. The lengths of these two images are set as 32. So, the three haptic images with dimensions of 32 × 32, 16 × 32, 8 × 32 are used for comparing the recognition performances of haptic images with different signal channels.

In summary, the process of generating haptic images is shown in Figure 9.

3.7 Train/test splits

The haptic images and the labels of 56 objects are partitioned into a training set and a testing set. The training set contains 4,600 haptic images of 46 objects, and the testing set contains 1,000 images of the other 10 objects. The partition is random, but both the training set and the testing set include at least one object in each category, and the two sets do not include the same object. The number of objects in each category is mainly decided by the proportion of this category in the total objects. The objects in the testing set are shown in Figure 10.

Figure 9 The process of generating haptic images



4. Haptic image recognition model

4.1 Model structure

A multi-class and multi-label model based on CNN is built for recognizing haptic features and object categories, as seen in Figure 11. The model is implemented by Caffe, which is an open-source deep learning framework (Jia *et al.*, 2014):

- *Haptic_Input* is a data layer. Its output has two parts: one is a label matrix, which is divided into five label vectors by *Slicer* layer, namely, the labels of hardness, thermal

Figure 10 The objects in the testing set

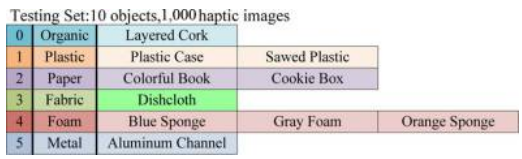
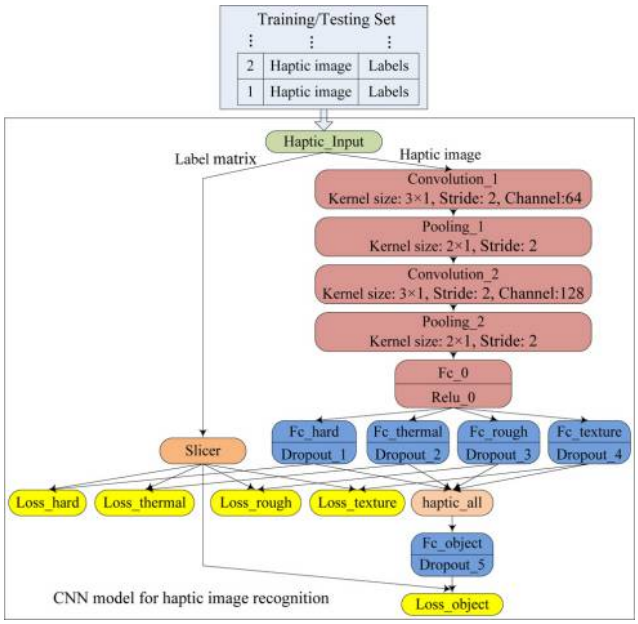


Figure 11 The CNN model for haptic image recognition



conductivity, roughness, texture and object categories (see Section 3.1); the other is the haptic images.

- *Convolution_1* and *Convolution_2* are convolutional layers. "Grouping" is used in the two convolutional layers, which makes different groups of channels not interact with each other. The number of groups is equal to the channel number of haptic images, which means interactions between different haptic signals are not learned until the fully connected layer (Gao *et al.*, 2016). This method is crucial for such a small training set because it can reduce the number of model parameters and improve the training efficiency.
- *Pooling_1* and *Pooling_2* are pooling layers with average pooling function.
- *Fc_0* is a fully connected layer with 64 neurons to fuse the features from the upper layer.
- *Relu_0* is an activation layer with Rectified Linear Units function, which has faster convergence rate than other functions such as *sigmoid* and *tanh* (Krizhevsky *et al.*, 2017).
- *Fc_hard*, *Fc_thermal*, *Fc_rough*, *Fc_texture* and *Fc_object* are fully connected layers and their outputs represent hardness, thermal conductivity, roughness, texture and object categories, respectively. In fact, *Fc_object* is the fusion of the haptic features from *Fc_hard*, *Fc_thermal*, *Fc_rough* and *Fc_texture*, namely, the category of an object is determined by the four extracted features. *Haptic_all* is provided for concatenating the output features from *Fc_hard*, *Fc_thermal*, *Fc_rough* and *Fc_texture*.
- *Dropout_1*, *Dropout_2*, *Dropout_3*, *Dropout_4* and *Dropout_5* are dropout layers. Dropout can reduce over-fitting by dropping out some nodes of the model with a certain probability.
- *Loss_hard*, *Loss_thermal*, *Loss_rough*, *Loss_texture* and *Loss_object* are loss layers with type of *SoftmaxWithLoss*, which is the combination of *softmax* function and multinomial logistic loss for multi-class classification tasks.

4.2 Model training

The proposed model is trained by the Stochastic Gradient Descent algorithm with a variable learning rate and a momentum value of 0.9. Learning rate can influence the speed and quality of the model training, as a high value may cause unstable oscillation, while a too low value slows the training. Therefore, it would better to update the learning rate during the training process:

$$lr(iter) = base\_lr \times (1 + \gamma \times iter)^{-\beta} \quad (2)$$

where $base\_lr = 0.01$, $\gamma = 0.0001$, $\beta = 0.75$ and *iter* is the number of iterations. Consequently, the learning rate *lr* will decrease according to *iter*. The maximum number of iterations is 600, and the batch size is 1,000.

5. Results

After training the model, the testing set is adopted to evaluate the recognition performance. The results are reported by F1-score, which reaches its best value at 100 per cent and worst value at 0 (Chu *et al.*, 2015). The following F1-score (per cent) is the weighted average of F1-scores of all classes for one label, where the weights are determined by the proportion of one class in the whole set.

5.1 Single-modal recognition

The single-modal haptic signals from four EPs are applied in haptic image recognition first. The haptic images with only PAC signals have four channels (one PAC in each EP), and their length is set as 32. Similarly, the haptic images with only PDCs, TDCs or TACs have a dimension of 4 × 32 each. The haptic image dimension for electrodes is 16 × 32 (four electrode signals after the PCA processing in each EP). The results are shown in Table II, Table III and Table IV, respectively.

As seen in Table II, Table III, Table IV:

5.1.1 For object recognition

The F1-score of PDCs is the lowest, because most of the objects in the same category have different hardness except

Table II Recognition with PACs (%)

Haptic images	Roughness	Texture	Object recognition
PACs (4 × 32)	67.4	61.4	64.2

Table III Recognition with PDCs, electrodes (%)

Haptic images	Hardness	Object recognition
PDCs (4 × 32)	46.9	51.4
Electrodes (16 × 32)	47.7	62.8

Table IV Recognition with TDCs, TACs (%)

Haptic images	Thermal conductivity	Object recognition
TDCs (4 × 32)	56.4	59.3
TACs (4 × 32)	64.0	68.6

metal objects, making it difficult to recognize objects by PDCs alone. Electrodes contain multi-dimensional features, resulting in obtaining a higher F1-score than PDCs. Objects can be more accurately recognized by extracting surface features from PACs. The F1-score of TACs is the highest, because the objects in PHAC-2 are classified by their materials, and the thermal conductivities of different materials are usually different leading to significant difference among the measured TACs. Although TDCs are able to recognize objects by different thermal changes, the thermal conduction of BioTac is very slow causing a little change of TDCs in a short time. Hence the TDCs' performance is much worse than that of TACs.

5.1.2 For feature extraction

As for thermal conductivity, the F1-score of TACs is higher than that of insensitive TDCs. Compared with PDCs, Electrodes have slightly better performance on hardness extraction, which may be due to their multi-dimensional features.

5.2 Multi-modal recognition

The five haptic images in Figure 8 are adopted to illustrate the recognition performances of multi-modal haptic images. The loss curves during the training with 32 × 32 haptic images are shown in Figure 12, which illustrate the convergence of the CNN model. The testing results are shown in Table V.

In Table V, the average F1-score of 32 × 32 haptic images is the highest, while decreasing the lengths (32 × 16 and 32 × 8) and the channels (8 × 32 and 16 × 32) of haptic images makes the average F1-score lower. The reasons should be that: shortening the lengths possibly discards some important information; less signal channels reduce the richness and diversity of haptic images. Therefore, a haptic image with more signal channels and longer length can have better recognition performance if there are enough haptic data to train the CNN model.

Figure 12 The loss curves during the training with 32 × 32 haptic images

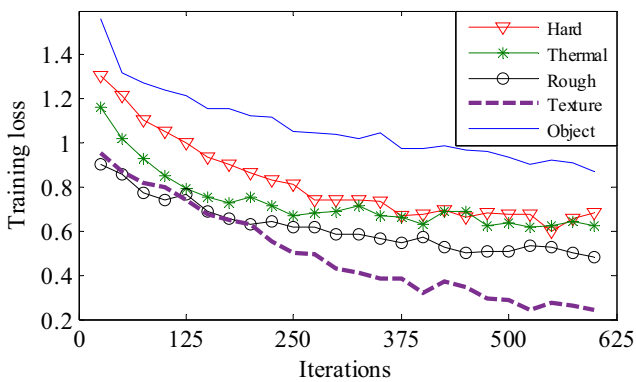


Table V Recognition with multi-modal haptic images (%)

Image dimensions	8 × 32	16 × 32	*32 × 32*	32 × 16	32 × 8
Hardness	58.6	58.8	*73.7*	55.9	57.0
Thermal conductivity	62.6	63.5	*67.7*	68.3	67.0
Roughness	53.8	68.3	*69.9*	68.9	69.4
Texture	53.5	68.8	*77.5*	75.6	67.4
Object recognition	60.5	69.0	*74.8*	67.6	67.2
Average	*57.8*	*65.7*	*72.7*	*67.3*	*65.6*

The confusion matrices of object recognition are shown in Figure 13. Apparently, *organic* objects are easily mistaken for *paper* objects, and *fabric* objects are easily mistaken for *foam* objects. There are two main reasons:

1. The number of objects in *organic* and *fabric* categories is too small to learn their features (five *organic* objects and seven *fabric* objects).
2. In the testing set (Figure 10), the *organic* object is *layered cork*, which is similar with some *paper* objects such as *fiber board*; the *fabric* object is *dishcloth*, which is similar with some *foam* objects such as *shelf liner*. These similarities also result in the wrong recognition.

However, the model can recognize *metal* objects rightly although the number of *metal* objects is smaller. It is because *metal* objects have some special features compared with other objects such as very high thermal conductivities.

The confusion matrices of hardness recognition are shown in Figure 14, which are used as an example for showing the haptic feature recognition performance. *hard-1* and *hard-2*, *hard-3* and *hard-4*, *hard-5* and *hard-6* are more easily confused. The main reason should be that the differences between these classes are not obvious enough. Furthermore, the measurement errors by BioTac can also lead to the misclassification.

6. Discussion

Compared with the results in Table II, Table III and Table IV, the performance of 32 × 32 haptic images in Table V is the best on both feature extraction and object recognition. It implies that multi-modal signal fusion has advantages over single-modal signals. Actually, the BioTac's output signals are not completely independent. For example, even for the same object, larger pressure (larger PDC) can lead to faster thermal conduction (larger TAC). Furthermore, the four features also have certain relations with each other. For example, harder objects often have higher thermal conductivities than softer objects, which means a haptic feature may be determined by more than one type of signals. Thus, fusing the multi-modal signals may be the only way to learn the complex coupling relations for more accurate recognition.

Other key factors, which influence the recognition performance, should be that: there is a small number of objects in some categories; the differences among different categories are not obvious. A direct method of solving these problems is to enlarge and enrich the haptic data set.

Different from the work by Chu *et al.* (2015) and Gao *et al.* (2016), the recognition model in this paper is not a binary classifier, which only outputs "yes" or "no", but a multi-class and multi-label classifier, which provides more accurate description of haptic features.

7. Conclusion

A new multi-class and multi-label deep learning model is designed for recognizing haptic images. The model can simultaneously extract four haptic features and recognize

Figure 13 Confusion matrices (per cent) of object recognition using the five haptic images: (a) is for 32 × 32 haptic images, (b) is for 32 × 16, (c) is for 32 × 8, (d) is for 8 × 32, (e) is for 16 × 32

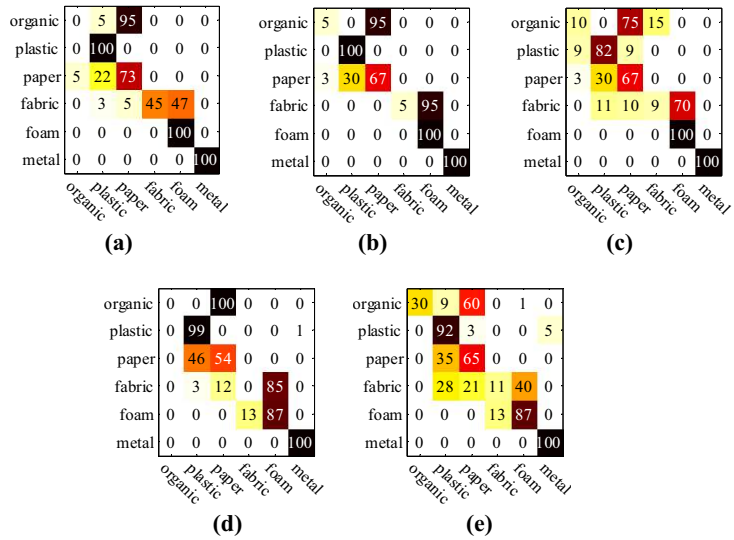
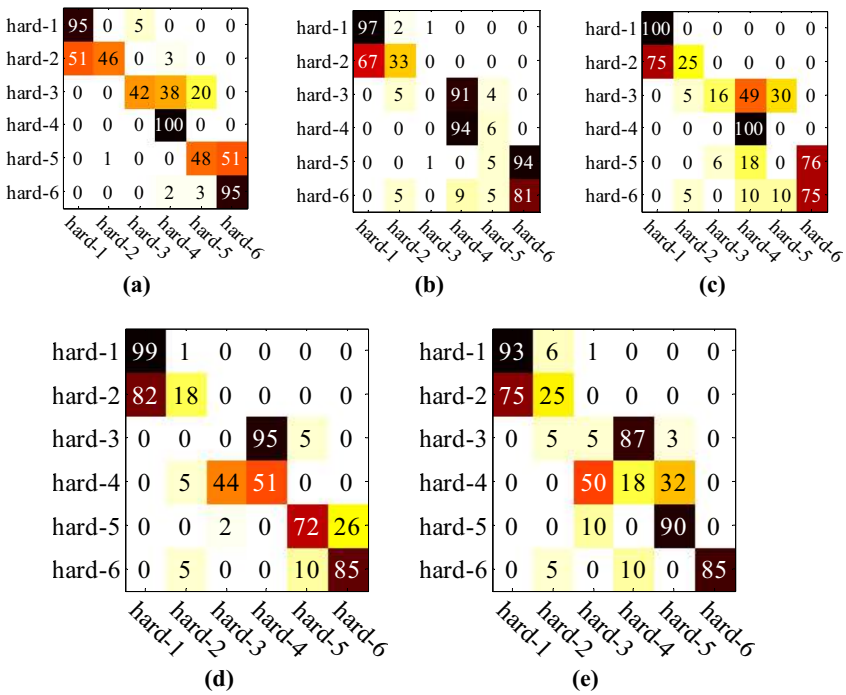


Figure 14 Confusion matrices (per cent) of hardness recognition using the five haptic images: (a) is for 32 × 32 haptic images, (b) is for 32 × 16, (c) is for 32 × 8, (d) is for 8 × 32, (e) is for 16 × 32



objects from the haptic images with multi-modal signals. The results show that: compared with single-modal haptic signals, multi-modal signal fusion can successfully improve the recognition performance; increasing the dimension of haptic images is also beneficial for better recognition performance. The proposed model can provide more accurate and richer haptic information, and improve the capabilities of unknown environment perception and dexterous manipulation.

References

Abdi, H. and Williams, L.J. (2010), "Principal component analysis", *Wiley interdisciplinary Reviews: Computational Statistics*, Vol. 2 No. 4, pp. 433-459.

Abdulnabi, A.H., Wang, G., Lu, J. and Jia, K. (2015), "Multi-task CNN model for attribute prediction", *IEEE Transactions on Multimedia*, Vol. 17 No. 11, pp. 1949-1959.

Arian, M.S., Blaine, C.A., Loeb, G.E. and Fishel, J.A. (2014), "Using the BioTac as a tumor localization tool", *IEEE Haptics Symposium*, IEEE, Houston, TX, 23-26 Feb., pp. 443-448.

Chen, G., Xia, Z., Sun, R., Wang, Z. and Sun, L. (2013), "A learning algorithm for model-based object detection", *Sensor Review*, Vol. 33 No. 1, pp. 25-39.

Chitta, S., Sturm, J., Piccoli, M. and Burgard, W. (2011), "Tactile sensing for mobile manipulation", *IEEE Transactions on Robotics*, Vol. 27 No. 3, pp. 558-568.

Chu, V., McMahon, I., Riano, L., McDonald, C.G., He, Q., Perez-Tejada, J.M., Arrigo, M., Darrell, T. and Kuchenbecker, K.J. (2015), "Robotic learning of haptic adjectives through physical interaction", *Robotics and Autonomous Systems*, Vol. 63 No. P3, pp. 279-292.

Chu, V., McMahon, I., Riano, L., McDonald, C.G., He, Q., Perez-Tejada, J.M., Arrigo, M., Fitter, N., Nappo, J.C., Darrell, T. and Kuchenbecker, K.J. (2013), "Using robotic exploratory procedures to learn the meaning of haptic adjectives", *IEEE International Conference on Robotics and Automation*, IEEE, Karlsruhe, 6-10 May, pp. 3048-3055.

Cu, C., Liu, W., Ruan, X. and Fu, X. (2016), "A Novel Tactile Sensing Array with Elastic Steel Frame Structure for Prosthetic Hand", *International Conference on Intelligent Robotics and Applications*, Springer, Tokyo, 22-24 August, pp. 471-478

Dargahi, J. and Najarian, S. (2004), "Theoretical and experimental analysis of a piezoelectric tactile sensor for use in endoscopic surgery", *Sensor Review*, Vol. 24 No. 1, pp. 74-83.

Fishel, J.A. (2012), "Design and use of a biomimetic tactile microvibration sensor with human-like sensitivity and its application in texture discrimination using Bayesian exploration", PhD Thesis, University of Southern California.

Fishel, J.A. and Loeb, G.E. (2012a), "Sensing tactile microvibrations with the BioTac— Comparison with human sensitivity", *IEEE Ras & Embs International Conference on Biomedical Robotics and Biomechatronics*, IEEE, Rome, 24-27 June, pp. 1122-1127.

Fishel, J.A. and Loeb, G.E. (2012b), "Bayesian exploration for intelligent identification of textures", *Frontiers in neurorobotics*, Vol. 6 No. 6, pp. 15-34.

Gao, Y., Hendricks, L.A., Kuchenbecker, K.J. and Darrell, T. (2016), "Deep learning for tactile understanding from visual and haptic data", *IEEE International Conference on Robotics and Automation*, IEEE, Stockholm, 16-21 May, pp. 536-543.

Gorges, N., Navarro, S.E., GöGer, D. and Wörn, H. (2010), "Haptic object recognition using passive joints and haptic key features", *IEEE International Conference on Robotics and Automation*, IEEE, Anchorage, AK, 3-7 May, pp. 2349-2355.

Han, D., Nie, H., Chen, M. and Wang, X. (2016), "Temperature prediction with humanoid finger sensor based on Fourier's law", *Measurement Science and Technology*, Vol. 27 No. 2, pp. 025105- 1-025105-7.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T. (2014), "Caffe: Convolutional architecture for fast feature embedding", *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, Orlando, FL, 3-7 Nov., pp. 675-678.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017), "Imagenet classification with deep convolutional neural networks", *Communications of the ACM*, Vol. 60 No. 6, pp. 84-90.

Lin, C.H., Erickson, T.W., Fishel, J.A., Wettels, N. and Loeb, G.E. (2009), "Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities", *IEEE International Conference on Robotics and Biomimetics*, IEEE, Guilin, 19-23 Dec., pp. 129-134.

Orii, H., Tsuji, S., Kouda, T. and Kohama, T. (2017), "Tactile texture recognition using convolutional neural networks for time-series data of pressure and 6-axis acceleration sensor", *IEEE International Conference on Industrial Technology*, IEEE, Toronto, ON, 22-25 March, pp. 1076-1080.

Song, A., Han, Y., Hu, H. and Li, J. (2014), "A novel texture sensor for fabric texture measurement and classification", *IEEE Transactions on Instrumentation and Measurement*, Vol. 63 No. 7, pp. 1739-1747.

Tian, L., Song, A. and Chen, D. (2017), "Image-based haptic display via a novel pen-shaped haptic device on touch screens", *Multimedia Tools and Applications*, Vol. 76 No. 13, pp. 1-24.

Tian, L., Song, A., Chen, D. and Ni, D. (2016), "Haptic display of image based on multi-feature extraction", *International Journal of Pattern Recognition & Artificial Intelligence*, Vol. 30 No. 8, pp. 1655023-1-1655023-24.

Wettels, N. and Loeb, G.E. (2011), "Haptic feature extraction from a biomimetic tactile sensor: Force, contact location and curvature", *IEEE International Conference on Robotics and Biomimetics*, IEEE, Karon Beach, 7-11 Dec., pp. 2471-2478.

Zhang, M.M. Atanasov, N. and Daniilidis, K. (2017), "Active tactile object recognition by monte carlo tree search", [online] Cornell University Library, available from: https://arxiv.org/abs/1703.00095[accessed 30 Jul 2017].

Corresponding author

Hong Nie can be contacted at: hnie@nuaa.edu.cn