Multi-modal broad learning for material recognition

Zhaoxin Wang1,2 | Huaping Liu1,2 | Xinying Xu1,2 | Fuchun Sun1,2

1Department of Computer Science and Technology, Tsinghua University, Beijing, China
2State Key Laboratory of Intelligent Technology and Systems, Beijing National Research Center for Information Science and Technology, Beijing, China

Abstract
Material recognition plays an important role in the interaction between robots and the external environment. For example, household service robots need to replace humans in the home environment to complete housework, so they need to interact with daily necessities and obtain their material performance. Images provide rich visual information about objects; however, it is often difficult to apply when objects are not visually distinct. In addition, tactile signals can be used to capture multiple characteristics of objects, such as texture, roughness, softness, and friction, which provides another crucial way for perception. How to effectively integrate multi-modal information is an urgent problem to be addressed. Therefore, a multi-modal material recognition framework CFBRL-KCCA for target recognition tasks is proposed in the paper. The preliminary features of each model are extracted by cascading broad learning, which is combined with the kernel canonical correlation learning, considering the differences among different models of heterogeneous data. Finally, the open dataset of household objects is evaluated. The results demonstrate that the proposed fusion algorithm provides an effective strategy for material recognition.

1 | INTRODUCTION

We are exposed to different types of materials every day and constantly evaluate the displayed features. For example, we can determine whether objects can be safely heated in a microwave oven by material recognition. On rainy days, we will walk on hard asphalt or slate road instead of muddy dirt road. Human beings can easily recognize object materials by prior knowledge and advanced perception system, but it is a challenge for robots. So, this research is highly desired for the development of computer visions [1].

To allow robots to observe, touch and knead the material properties of objects as humans do, a variety of sensors are fitted in them. Abundant visual information of objects is provided by cameras, which becomes one of the most commonly used sensors in the field of artificial intelligence [2]. However, it is challenging for the robot to determine the material of the object visually. For example, choose a ceramic cup from a red plastic cup and a red ceramic cup. So, it’s not always possible to recognize objects from visual cues and infer their properties [3]. Because humans have advanced tactile sensing systems, ceramic cups can be selected by touch. Therefore, tactile sensors are applied in many research works related with robots, including terrain classification [4], object location [5], force measurement of stable grasping [6], object instance recognition [7], control of unknown objects [8], and so on. Compared with visual signals, tactile signals not only process high sensitivity but also can be directly detected by a variety of characteristics of the measured object, such as material, texture, and temperature [9], contact vibration [10], and shape structure [11]. Detailed tactile perception of robots can be referred to Luo et al. [12]. For material classification, Fu et al. regarded tactile sound as a model of wood material attribute perception [13]. By tapping on different kinds of woodblocks, the sound was collected and analysed to obtain the material properties of different kinds of wood. The ground classification based on ontology perception mainly makes use of the mechanical characteristics of vibration signals [14, 15] and tactile sensation [16] to complete the identification and classification of the terrain.

Although the application of tactile modal and visual modal information has obtained many significant achievements in the field of material recognition, either single visual data or tactile data is limited in the expressions of the material characteristics. Therefore, many researchers have adopted multi-modal method. Zheng et al. [17] used tactile acceleration signals and corresponding surface texture images to classify surface materials. Liu et al. [2] established a projection dictionary.
learning framework for weakly paired multi-modal data fusion, indicating the effectiveness of the algorithm on 53 household objects containing visual and tactile information. In addition to the visual–tactile fusion, the accuracy in material recognition can be improved by two kinds of touch models of data fusion. By the method of fusion of tactile sense and depth camera, a six-legged walking robot [16] could identify 12 surfaces with an accuracy of 95%.

With the rapid development of material recognition and deep learning, many achievements are obtained by the multi-modal material recognition method of automatic feature extraction. However, owing to long training time of deep learning, and it is easy to have the problem of local optima, difficult to converge. The paper aims to search a simple and efficient machine learning method, which can reduce the training time of the model, save costs, and improve the accuracy of material recognition. Broad learning (BRL) was first invented by Chen and Liu [18], which has attracted a lot of academic attention due to its rapid incremental learning and alternatives for deep learning. Therefore, many researchers have improved BRL according to the application background of specific problems [19–23], and it can be applied in the fields of image recognition [24], classification and regression [25], and data modelling [26]. Based on the feedforward neural network of a single-layer random vector functional link neural network (RVFLNN), the original input is firstly learned from the sparse mapping features through the feature nodes, and then the enhanced features are achieved through the non-linear expansion of the enhancement nodes. The parallel expressions of the two features are sent to the output layer as the final total output for recognition. Therefore, this method not only can obtain better flexibility and generalization performance, but also avoids many problems of deep learning. However, it is difficult to capture the relevant high-level abstract features to complete the material recognition task of multi-modal data owning to the single-layer feedforward neural network, which is designed to use linear mapping to form feature nodes.

In this paper, a BRL algorithm based on the multi-modal framework is proposed, which can efficiently integrate various forms of information for material recognition. The architecture is not only robust for diverse visual and tactile signals but also makes full use of multi-modal information to extract more abstract representations and invariant features. We conducted experiments on the MREO (material recognition of everyday objects) dataset.

Herein, the main contributions are summarized as follows:

- In order to improve the material recognition performance and reflection ability of dexterous robot, the algorithm CFBRL-CCA in this paper can learn the non-linear features between multiple modes and obtain satisfactory results in a relatively short time.
- When dealing with heterogeneous multi-modal data collected by different types of sensors, a cascading broad learning algorithm for material recognition is proposed. First, two cascading BRL subnetworks are used to extract the feature matrix of different modal data, respectively. Then, correlation learning is carried out by using the kernel canonical correlation analysis (KCCA). Finally, the recognition results are obtained by the BRL classifier.
- CFBRL-KCCA is evaluated on MREO dataset. The experimental results show it shortens the training time, improves the correlation between different modal data, and efficiently completes the material recognition task.

## 2 PRELIMINARIES

The recently proposed rapid incremental learning network-BRL [18] is designed based on the idea of RVFLNN, which provides an alternative method for deep learning. The structure of the system is shown in Figure 1. The whole network is composed of four parts: input, feature nodes, enhancement nodes, and output.

The basic structure and calculation steps of BLS are as follows. The given input data is $X \in \mathbb{R}^{B \times C}$, and output is $Y \in \mathbb{R}^{B \times D}$. First, the $m$-group feature map is generated, and each group contains $f_i$ feature nodes. Then, the mapping feature of the $i$th group is expressed as follows:

$$F_i = G(X \cdot W_{f_i} + \beta_{f_i}), i = 1, 2, \ldots, m$$ (1)

where $G$ is a linear activation function, which makes the mapping linear feature. In the process of complex input data, it is not enough to extract many useful features. Therefore, this gives researchers the idea of improving BRL. $W_{f_i}$ and $\beta_{f_i}$ are the weights and biases of random initialization. In order to overcome the unpredictability of random initialization, BRL adopts the sparse autocoecoding to optimize the input weights.

Then, $F = [F_1, F_2, \ldots, F_m]$ is defined as the series mapping feature of group $i$, and the enhancement nodes of group $j$ are expressed as

$$E_j = \xi(F^m \cdot W_{ej} + \beta_{ej}), j = 1, 2, \ldots, M$$ (2)

where $\xi$ is a non-linear activation function. Tanh function is used here. $W_{ej}$ and $\beta_{ej}$ are the weights and biases of enhancement nodes. Similarly, enhancement nodes of $M$ group in series are defined as $E^M = [E_1, E_2, \ldots, E_M]$.

Extract sparse features from the given training data, consider using sparse coding to solve input optimization problems:

$$\arg \min \|Y \hat{W} - X\|^2_2 + \lambda \|\hat{W}\|^1_1.$$ (3)

where $\hat{W}$ is the sparse self-coding solution. $Y$ is the expected output of a given linear equation, that is $X \cdot W = Y$. The above problem is expressed as a convex problem in [27] by lasso. Therefore, the approximation problem can be solved in many ways, such as K-Singular Value Decomposition (SVD), Alternating Direction Method of Multipliers (ADMM), and Fast Iterative Shrinkage Thresholding Algorithm. Among them, the
ADMM method is actually designed for general decomposition methods and decentralized algorithms in optimization problems. Equation (3) can be equivalently regarded as the following general problem:

\[
\arg \min_{W} f(W) + g(\theta), \text{ s.t. } W - \theta = 0. \tag{4}
\]

Therefore, the proximal problem can be solved by the following iterative steps:

\[
\begin{aligned}
W_{k+1} &= (Z^T Z + \rho I)^{-1}(Z^T X + \rho(\theta_k - u_k)) \\
o_{k+1} &= S_\rho(\theta_k) (W_{k+1} + u_k) \\
u_{k+1} &= u_k + (W_{k+1} - o_{k+1})
\end{aligned}
\tag{5}
\]

where \(\rho > 0\), \(S\) is a soft threshold operator, defined as

\[
S_\rho(a) = \begin{cases} 
\frac{a - k}{\rho}, & a > k \\
0, & |a| \leq k \\
\frac{a + k}{\rho}, & a < -k
\end{cases}
\tag{6}
\]

The output of the BRI model can be expressed as follows:

\[
Y = [F_1, F_2, \ldots, F_m] E_1, E_2, \ldots, E_M W^M
\]
\[
= [E^M] W^M
\]
\[
= [A_m^M] W^M
\tag{7}
\]

In the planar network, pseudo-inverse is a very convenient method to solve the weight of the output layer of the neural network. In fact, the generalized inverse can be calculated in different ways, such as orthogonalization method, orthogonal projection method, SVD, and iterative method. However, the direct solution is too time consuming and labour intensive, especially the training samples and input modalities are characterized by large scale and variability [28]. In fact, the following optimal problem is another way to solve the pseudo-inverse:

\[
\arg \min_{W} \|AW - Y\|_F^2 + \lambda \|W\|_F^2 \tag{8}
\]

The above optimal problem is set to regular \(L_2\) regularization by \(\alpha_1 = \alpha_2 = \mu = \nu = 2\). It has good generalization performance. The solution of \(L_2\) regularization is equivalent to the ridge regression theory. That is,

\[
W^M = [A_m^M]^+ Y \tag{9}
\]

\([A_m^M]^+\) is obtained by the following formula:

\[
[A_m^M]^+ = \lim_{\lambda \to 0} (\lambda I + A_m^M \cdot [A_m^M]^T)^{-1} [A_m^M]^T \tag{10}
\]

If the test accuracy does not meet the requirements or new data is entered, the network needs to be expanded using an incremental learning algorithm without a retraining process.

## 3 | ARCHITECTURE

The recently proposed rapid incremental learning network-BRI [18] is designed based on the idea of RVFLNN, which provides an alternative method for deep learning. The whole
network is composed of four parts: input, feature nodes, enhancement nodes, and output. CFBRL-KCCA is proposed based on the original BRL.

According to the difference of input data types, multimodal data can usually be divided into homogeneous data and heterogeneous data. For the latter, it is necessary to mine the potential non-linear relationship between different modal for material recognition. However, each modal data has its most important features, such as vibration and sound signals; their powerful features are the variation rule. So, vibration signals are usually converted to the spectrum by the short-time Fourier transform or the Mel bank features. For image features, algorithm mainly learn their colour and texture features. To this end, we propose a multi-modal data fusion framework. First, the features of two modal data are extracted through CFBRL [22], respectively. Then KCCA is used to maximize the correlation between the two modal data. Finally, the classification results are obtained by the BRL classifier. The framework is shown in Figure 2, which consists of three main parts:

1) For the two modal data after pre-processing, unsupervised learning of feature representation of cascading feature nodes and enhancement nodes is carried out by the CFBRL algorithm. The given input data is \( X \in \mathbb{R}^{B \times C} \), \( Y \in \mathbb{R}^{B \times D} \). The first group of feature nodes is expressed as

\[
F_{c1} = \phi(X \cdot W_{cf1} + \beta_{cf1})
\]

where \( \phi \) is a linear activation function, \( W_{cf1} \) and \( \beta_{cf1} \) are randomly generated weights and bias with corresponding dimensions. Then generate a second set of mapped features, and use the output \( F_{c1} \) of the first set of feature nodes to establish a mapped feature \( F_{c2} \):

\[
F_{c2} = \phi(F_{c1} \cdot W_{cf2} + \beta_{cf2}) \\
= \phi(\phi(X \cdot W_{cf1} + \beta_{cf1}) \cdot W_{cf2} + \beta_{cf2}) \\
= \phi^2(X; \{W_{cf1}; \beta_{cf1}\}_{i=1,2})
\]

The \( n \)th set of feature nodes is expressed as

\[
F_{cn} = \phi(F_{cn-1} \cdot W_{cfn} + \beta_{cfn}) \\
= \phi^i(X; \{W_{cf1}; \beta_{cf1}\}_{i=1})
\]

Then, \( F^p = [F_{c1}, F_{c2}, \ldots, F_{cm}] \) is defined as the series mapping feature group. It used to generate enhancement nodes to extract high-level features of data, and the enhancement nodes of the \( j \)th group are expressed as

\[
E_{cj} = \xi(F^p \cdot W_{eqj} + \beta_{eqj}), j = 1, 2, \ldots, m
\]

where \( \xi \) is a non-linear activation function. Tanh function is used here. \( W_{eqj} \) and \( \beta_{eqj} \) are the weights and biases of enhancement nodes. Similarly, enhancement nodes of \( M \) group in series are defined as \( E^m = [E_{c1}, E_{c2}, \ldots, E_{cm}] \). The above \( F^p \) and \( E^m \) form the preliminary feature matrix \( H_c \) of the first cascading BRL subnetwork. \( H_c \) is expressed as

\[
H_c = [F_{c1}, F_{c2}, \ldots, F_{cm} | E_{c1}, E_{c2}, \ldots, E_{cm}]
\]

Preliminary feature matrix \( H_c \) of the second cascading BRL subnetwork is represented as

\[
H_s = [F_{c1}, F_{c2}, \ldots, F_{CM} | E_{c1}, E_{c2}, \ldots, E_{sM}]
\]
2) KCCA is used to achieve joint feature dimension reduction by maximizing the correlation between two modal features.

KCCA can reveal the potential non-linear relationship between variables, map the samples of the original space to the high-dimensional feature space with kernel techniques. It extracts the characteristics of variables in the feature space, and thus solves the non-linear problem of the original space.

Suppose the sample matrix of modal one and modal two are \( H_c = (X_1, X_2, \ldots, X_n) \) and \( H_s = (Y_1, Y_2, \ldots, Y_n) \), respectively. \( \Phi_{H_c}(X) \) and \( \Phi_{H_s}(Y) \) represent the transformation of nonlinear mapping \( \Phi(\cdot) \) on \( H_c \) and \( H_s \), respectively (the mapping from \( s, t \) dimension to \( n \) dimension). That is,

\[
\Phi_{H_c}(X) = (\Phi_{H_c}(X_1), \Phi_{H_c}(X_2), \ldots, \Phi_{H_c}(X_n)) \\
\Phi_{H_s}(Y) = (\Phi_{H_s}(Y_1), \Phi_{H_s}(Y_2), \ldots, \Phi_{H_s}(Y_n))
\] (17)

The transformed \( \Phi_{H_c}(X) \) and \( \Phi_{H_s}(Y) \) are both \( n \times N \)-dimensional matrices. In the feature space, canonical correlation analysis (CCA) is carried out on the mapped samples.

Like CCA, two \( n \)-dimensional vectors \( \alpha = [a_1, a_2, \ldots, a_N]^T \) and \( \beta = [\beta_1, \beta_2, \ldots, \beta_N]^T \) are found in high-dimensional space, so that the linear combination of the two samples, \( H_{cc} = \Phi_{H_c}(X) \cdot \alpha \) and \( H_{ss} = \Phi_{H_s}(Y) \cdot \beta \), has the largest phase coefficient, \( \rho \). The solution process of vectors \( \alpha \) and \( \beta \) is transformed into the following constrained optimization problem:

\[
\max \rho(H_{cc}, H_{ss}) = \alpha^T \Sigma H_{cc} H_{ss} \beta
\] (18)

where \( \Sigma H_{cc} H_{ss} \) is the cross-covariance matrix of \( H_c \) and \( H_s \).

Then, the objective function of KCCA is

\[
H_{cc}^T \Phi_{Hc}(X) \Phi_{Hs}(Y)^T H_{ss} = (\Phi_{Hc}(X) \alpha)^T \Phi_{Hs}(Y) \beta
\] (19)

The kernel matrices are defined on \( H_c \) and \( H_s \) as follows:

\[
K_{Hc} = \Phi_{Hc}(X)^T \Phi_{Hc}(X) \\
K_{Hs} = \Phi_{Hs}(Y)^T \Phi_{Hs}(Y)
\] (20)

Rewrite the objective function as

\[
\max \rho(H_{cc}, H_{ss}) = \alpha^T K_{Hc} K_{Hs} \beta
\] (21)

Suppose the constraints are

\[
\begin{align*}
\alpha^T K_{Hc} K_{Hs} \alpha &= \beta^T K_{Hs} K_{Hc} \beta = 1 \\
H_{cc} &= \Phi_{Hc}(X) \alpha \\
H_{ss} &= \Phi_{Hs}(Y) \beta
\end{align*}
\] (22)

Introducing the Lagrange multiplier, using Equations (21) and (22), the following characteristic problems are obtained:

\[
\begin{bmatrix}
0 & K_{Hc} K_{Hs} \\
K_{Hs} K_{Hc} & 0
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix} = \eta
\begin{bmatrix}
K_{Hc} K_{Hc} & 0 \\
0 & K_{Hs} K_{Hs}
\end{bmatrix}
\begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}
\] (23)

where \( \eta = \alpha^T K_{Hc} K_{Hs} \beta \) and then \( \text{alpha and beta are obtained.} \)

3) Connect the \( H_c \) and \( H_s \) in series, which is input for BRL. Recognition results are obtained. The training of the material recognition network is completed.

4 | EXPERIMENTAL RESULTS

4.1 | MREO dataset

The MREO dataset [29] used contains common objects in household environment, such as cups, bowls, pots etc. It is distributed across six material categories: plastic, glass, cloth, metal, wood, and ceramic. Each material has 12 different objects. So, there are 72 objects in the dataset. Vibrations, forces, and rate of heat transfer are obtained by contact microphones, two force sensors, and thermistor, respectively. In this experiment, vibration signal and force signal commonly used in human–computer interaction of robot are selected.

As can be seen from Figure 3, the cloth is easily distinguished from other materials by only using vibration signals. The main reason is that these soft clothes have no obvious vibration contact information. However, if the interaction time between the robot and the object is increased, a long period of sliding interaction can provide a more obvious vibration for the cloth. In this experiment, the vibration signals used are obtained when the object contacted the sensor for 0.2 s, and the force is obtained after the object contacted the sensor for 2 s. On the other hand, the other five materials will produce clearer vibration signals when the terminal controller of the robot contacts with the objects. To get appropriate vibration features, they are often transformed into the Mel-scaled spectrogram through Mel filter banks. They are shown in Figure 4.

Force signals are obtained by touching objects with two force sensors at the robot's fingertips. As can be seen from Figure 3, a shallow fusion of these two groups of data is constructed before extracting the features. Because characteristics of change rate or slope usually are used, it can be directly connected as the input of network.

4.2 | EXPERIMENTAL RESULTS AND ANALYSIS

The multi-modal fusion method integrates data according to the optimal principle to realize the complementary advantages
of multi-modal data. Therefore, effective methods for achieving the maximum correlation between multi-modal data need to be found. In this experiment, the eight most advanced algorithms of CCA, KCCA, BRL-CCA, BRL-KCCA, C-CFBRL-CCA, C-CFBRI-KCCA, DCCA, and DCCAE are compared from the aspects of correlation coefficient and recognition task. Among them, CCA [30] is a multi-dimensional statistical learning method. By maximizing the correlation of two sets of data in the same linear projection space, typical representative variables corresponding to two sets of data are obtained, respectively, to replace the original features to achieve high-dimensional feature fusion of multi-modal information. DCCA [31] is a method to learn the complex non-linear transformation of two sets of data. The flexible non-linear correlation representation of multi-modal data is learned through a deep neural network. DCCAE [32] is a new model composed of two auto-encoders, which optimizes the combination of the typical correlation between learning representation and auto-encoder reconstruction errors.

The relevance of the characteristics learned by these algorithms is compared firstly. It can be seen from Figure 5 that, with the increase of output dimension, the sum of correlation coefficients of the eight algorithms shows an upward trend. Random variables are linearly mapped by CCA, and the sum of correlation coefficients is the lowest. KCCA uses the kernel method for non-linear learning of variables, which can fully mine the nonlinear characteristics of two sets of data.

The sum of correlation coefficients obtained by the algorithm based on deep neural network is obviously superior to linear CCA and KCCA, and the advantage is more obvious when the output dimension is about 150. This is because the non-linear representation can more fully mine the similarity of two sets of data and obtain features with a higher correlation. However, in the training process of these two deep learning methods, the optimal parameters need to be obtained through the gradient descent method and multiple iterations until the error converges. The computational complexity is large, and it is easy to have the problem of local optima. The training time can be seen in Table 1.
The recognition accuracy using CCA is lower than that using KCCA. The main reason is that CCA seeks basis vectors for two sets of variables so that the correlation of the projections of these variables on the basis vectors is maximized. From the perspective of information theory, this transformation maximizes the mutual information between feature extraction. However, if there is a strong non-linear relationship between two sets of variables, CCA may not extract more useful data description features due to its linear mapping. The kernel technique of CCA can reveal the potential non-linear relationship among internal variables. In this experiment, the Gaussian kernel function is used.

By comparing the training time of several algorithms, it can be concluded that the multi-modal fusion algorithm based on BRL has shorter training time and better effect than DCCA and DCCAE. Deep learning training requires a lot of data, and its complex structure and long training time make it difficult to apply to the practical application of robots. BRL solves this problem. In the case of a small amount of data, the CFBRL-KCCA proposed herein achieves the classification accuracy of 87.89% with its simplified structure and good generalization performance. The training time is only 1/32 of that of DCCAE.

Figure 6 compares the classification accuracy of four BRL algorithms in single modal and multi-modal. On the whole, the recognition rate of materials using two modals is higher than that of single modal, which is because the features of different models can complement each other to complete the task of material recognition better, which also indicates the necessity of multi-modal feature fusion. On the other hand, using force modal data makes it easier to distinguish material than using vibration signals. This may be because the features of force signal are obvious, and the algorithm mainly learns its slope and variation trend features, so it has much less noise interference than the vibration sound signal. But the accuracy of combination of the two modals is still the highest.

Figure 7 is the classification results of the four algorithms in different output dimensions. The comparison between CCA and KCCA shows that the kernel method can better learn the nonlinear characteristics of the two modals and obtain a better recognition rate. When the output dimension of CCA and

![Figure 5](image)  
**Figure 5** The sum of the correlation coefficients of the first 50 typical variables

| Method     | Accuracy (%) | Training time (s) |
|------------|--------------|-------------------|
| CCA        | 68.57        | 10.41             |
| KCCA       | 78.83        | 17.28             |
| DCCA       | 82.55        | 585.66            |
| DCCAE      | 83.76        | 601.34            |
| BRL-CCA    | 84.23        | 10.92             |
| BRL-KCCA   | 85.88        | 17.55             |
| CFBRL-CCA  | 85.64        | 11.74             |
| CFBRL-KCCA | 87.89        | 18.03             |

![Figure 6](image)  
**Figure 6** Single-modal and multi-modal classification accuracy of different algorithms

![Figure 7](image)  
**Figure 7** Recognition accuracy of four algorithms under different output dimensions
KCCA is low, the recognition accuracy of the four algorithms is very close, about 82%. It can be seen from the Figure 7, subsequently, the recognition accuracy increases significantly with the increase of output dimension, and remains flat or decreased slightly after reaching the maximum value. The two algorithms proposed always show good effect, and achieve the maximum value when the output dimensions are 200 and 240, respectively.

5 | CONCLUSIONS

Herein, we apply BRL to the field of material recognition and propose a multi-modal material recognition framework based on CFBRL. This method combines unsupervised feature extraction, maximizing the correlation between two features to achieve joint feature dimensionality reduction, and using the BRL classifier to obtain material recognition results into a complete broad structure for material recognition tasks. Experimental results show that the proposed framework and method are effective in multi-modal material recognition.

ACKNOWLEDGEMENTS

This work was supported by Joint Fund of Science & Technology Department of Liaoning Province and State Key Laboratory of Robotics, China (2020-KF-22-06).

REFERENCES

1. Bell, S., et al.: Material recognition in the wild with the materials in context database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3479–3487. (2015)
2. Liu, H., et al.: Weakly paired multimodal fusion for object recognition. IEEE Trans. Autom. Sci. Eng. 15(2), 784–795 (2017)
3. Fishel, J.A., Loeb, G.E.: Bayesian exploration for intelligent identification of textures. Frontiers Neurorobot. 4 (2012)
4. Hsiao, K., Kaelbling, L., Lazor, T.: Task-driven tactile exploration. Frontiers Neurorobot. 6, 4 (2012)
5. Petrovskaya, A., Khatib, O.: Global localization of objects via touch. IEEE Trans. Robot. 27(3), 569–585 (2011)
6. Nguyen, K.C., Perdereau, V.: Fingertip force control based on max torque adjustment for dexterous manipulation of an anthropomorphic hand. In: 2013 IEEE/RSJ international Conference on Intelligent Robots and Systems, pp. 3557–3563. IEEE (2013)
7. Madry, M., et al.: 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)
8. Hallerberg, M.: Explaining European patterns of taxation: From the introduction of the euro to the euro-crisis. IDB Book No. IDB-TN-400 (2012)
9. Fishel, J.A., Loeb, G.E.: Bayesian exploration for intelligent identification of textures. Frontiers Neurorobot. 6, 4 (2012)
10. Goger, D., Gorges, N., Worn, H.: Tactile sensing for an anthropomorphic robotic hand: Hardware and signal processing. In: IEEE International Conference on Robotics and Automation, pp. 895–901. IEEE (2009)
11. Martinez-Hernandez, U., et al.: Active contour following to explore object shape with robot touch. In: 2013 World Haptics Conference (WHC), pp. 341–346. IEEE (2013)
12. Luo, S., et al.: Robotic tactile perception of object properties: A review. Mechatronics. 48, 54–67 (2017)
13. Fu, H., et al.: Wood material recognition for industrial applications. Syst. Sci. Control Eng. 6(3), 346–358 (2018)
14. Brooks, C.A., Jagmenn, K.: Vibration-based terrain classification for planetary exploration rovers. IEEE Trans. Robot. 21(6), 1185–1191 (2005)
15. Dupon, E.M., et al.: Frequency response method for terrain classification in autonomous ground vehicles. Auton. Robot. 24(4), 337–347 (2008)
16. Wang, Z., et al.: Deep learning for surface material classification using haptic and visual information. IEEE Trans. Multimedia, 18(12), 2407–2416 (2016)
17. Zheng, H., et al.: Deep learning for surface material classification using haptic and visual information. IEEE Trans. Multimedia, 18(12), 2407–2416 (2016)
18. Chen, C.L.P., Liu, Z.: Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. IEEE Trans. Neural Netw. Learn. Syst. 29(1), 10–24 (2017)
19. Li, Z., Chen, C.L.P.: Broad learning system: Feature extraction based on K-means clustering algorithm. In: 2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS), pp. 683–687. IEEE (2017)
20. Liu, Z., Chen, C.L.P.: Broad learning system: Structural extensions on single-layer and multi-layer neural networks. In: 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), pp. 136–141. IEEE (2017)
21. Jia, J., Liu, Z., Chen, C.L.P.: Discriminative graph regularized broad learning system for image recognition. Sci. China Inf. Sci. 61(11), 112209 (2018)
22. Chen, C.L.P., Liu, Z., Feng, S.: Universal approximation capability of broad learning system and its structural variations. IEEE Trans. Neural Netw. Learn. Syst. 30(4), 1191–1204 (2018)
23. Li, D., Shujuan, J., Chunjin, Z.: Improved broad learning system: Partial weights modification based on BP algorithm. In: IOP Conference Series: Materials Science and Engineering, vol. 439. IOP Publishing (2018)
24. Kong, Y., et al.: Hyperspectral imagery classification based on semi-supervised broad learning system. Remote Sens. 10(5), 685 (2018)
25. Feng, S., Chen, C.L.P.: Fuzzy broad learning system: A novel neuro-fuzzy model for regression and classification. IEEE Trans. Cybernetics. 50(2), 414–424 (2018)
26. Jia, J.-W., Philip, C.C.L.: Regularized robust broad learning system for uncertain data modeling. Neurocomputing. 322, 58–69 (2018)
27. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Series B. 58(1), 267-288 (1996)
28. Cohen, C.: Random Matrices: Theory and Applications (1986)
29. Ericksen, Z., Chernova, S., Kemp, C.C.: Semi-supervised haptic material recognition for robots using generative adversarial networks. arXiv preprint. arXiv:1707.02796 (2017)
30. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: An overview with application to learning methods. Neural Comput. 16(12), 2639–2664 (2004)
31. Andrew, G., et al.: Deep canonical correlation analysis: An overview with application to learning methods. Neural Comput. 16(12), 2639–2664 (2004)
32. Wang, W., et al.: On deep multi-view representation learning. In: International Conference on Machine Learning, pp. 1247–1255 (2013)
33. Wang, Z., et al.: Multi-Immodal broad learning for material recognition. Cogn. Comput. Syst. 3(2), 123–130 (2021).

How to cite this article: Wang, Z., et al.: Multi-Immodal broad learning for material recognition. Cogn. Comput. Syst. 3(2), 123–130 (2021). https://doi.org/10.1049/ccs.2.12004