SCIENCE CHINA
Information Sciences

• LETTER •

February 2023, Vol. 66 129201:1–129201:2
https://doi.org/10.1007/s11432-021-3512-6

Robotic haptic adjective perception based on coupled sparse coding

Pengwen XIONG[1*], Kongfei HE[1], Aiguo SONG[2*] & Peter X. LIU[3]

[1]*School of Advanced Manufacturing, Nanchang University, Nanchang* 330031*, China;*
[2]*School of Instrument Science and Engineering, Southeast University, Nanjing* 210096*, China;*
[3]*Department of Systems and Computer Engineering, Carleton University, Ottawa* ON KIS 5B6*, Canada*

Received 10 March 2021/Revised 25 November 2021/Accepted 14 May 2022/Published online 5 January 2023

Citation Xiong P W, He K F, Song A G, et al. Robotic haptic adjective perception based on coupled sparse coding. Sci China Inf Sci, 2023, 66(2): 129201, https://doi.org/10.1007/s11432-021-3512-6

Objects, textures, and materials can be identified by extracting haptic interaction information [1,2]. For haptic adjective understanding, Gao et al. [3] adopted deep models as a unified way to learn information from vision and haptics modalities. Nevertheless, this approach depends on a transfer learning method designed for object classification. Unlike feature learning methods, Chu et al. [4] designed a hand-crafted feature, which is hand-designed and usually has a clear physical meaning, to distinguish haptic adjectives. In general, most scholars attempt to solve the abstract and tortuous haptic problems with an entirely single-feature extraction method: a hand-crafted feature or a learning feature. In this study, we attempt to fuse measurements to avoid these defects based on a new sparse coding model. Sparse coding [5] is the representation of the back neurons by the strong activation of a relatively small group of preceding neurons and is widely used in unconventional unsupervised learning methods. Sparse coding is a remarkable tool for multimeasurement fusion that aims to find a complete sparse representation of the input data. Sparse coding tries to solve the primary problem:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y}-\boldsymbol{D}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1, \tag{1}$$

where $\boldsymbol{y}\in\mathbb{R}^{m\times 1}$ is the original data, $\boldsymbol{D}\in\mathbb{R}^{m\times n}$ is the dictionary matrix, $\boldsymbol{x}\in\mathbb{R}^{n\times 1}$ is the sparse coefficient vector, λ is a hyperparameter, $m<n$, and $\|\,\|_1$ is the sum of absolute values of elements in the vector $\boldsymbol{x}$.

We divide haptic data such as in Figure 1 into scalar and electrode array signals. Formally, the matrix $\boldsymbol{S}_G\in\mathbb{R}^{d_S\times N}$ constituted by vectors comprising the features of scalar signals is derived as

$$\boldsymbol{S}_G=\left[\boldsymbol{S}_{1,1},\boldsymbol{S}_{2,1},\ldots,\boldsymbol{S}_{N_1,1},\ldots,\boldsymbol{S}_{N_2,2},\ldots,\boldsymbol{S}_{N_o,O}\right],$$

where G denotes the different haptic adjectives and N denotes the number of samples in the training set. In the same way, the matrix $\boldsymbol{E}_G\in\mathbb{R}^{d_E\times N}$ constituted by vectors consisting of the features of electrode array signals is defined as

$$\boldsymbol{E}_G=\left[\boldsymbol{E}_{1,1},\boldsymbol{E}_{2,1},\ldots,\boldsymbol{E}_{N_1,1},\ldots,\boldsymbol{E}_{N_2,2},\ldots,\boldsymbol{E}_{N_o,O}\right].$$

The training set obtains $N=\sum N_i$ samples, where o indicates the last object and N_i indicates the number of samples of the i-th object. During the testing phase, all measurements of the test sample are simultaneously entered into the constructed classifier to obtain a common prediction label that lies in {positive, negative}.

Generally, for multisensor information fusion, a sparse coding method is used to attempt to solve the following optimization problem:

$$\min_{\boldsymbol{D}_S,\boldsymbol{D}_E,\boldsymbol{X}_S,\boldsymbol{X}_E}\Phi_R\left(\boldsymbol{D}_S,\boldsymbol{D}_E,\boldsymbol{X}_S,\boldsymbol{X}_E\right)+\Phi_P\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right), \tag{2}$$

where $\Phi_R\left(\boldsymbol{D}_S,\boldsymbol{D}_E,\boldsymbol{X}_S,\boldsymbol{X}_E\right)=\|\boldsymbol{S}_G-\boldsymbol{D}_S\boldsymbol{X}_S\|_F^2+\|\boldsymbol{E}_G-\boldsymbol{D}_E\boldsymbol{X}_E\|_F^2$, $\Phi_P\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right)$ is the penalty term, which contains a traditional sparse term and other penalty terms in many studies. $\boldsymbol{D}_S\in\mathbb{R}^{d_S\times K}$ is the dictionary matrix for scalar signals, and $\boldsymbol{D}_E\in\mathbb{R}^{d_E\times \boldsymbol{K}}$ is the dictionary matrix for electrode array signals. $\boldsymbol{X}_S$ is the matrix comprising sparse coefficient vectors of scalar signals in training samples, and $\boldsymbol{X}_E$ is the matrix comprising sparse coefficient vectors of electrode array signals in training samples.

We use two projection matrices that can perfectly project raw features to a higher dimensional space:

$$\boldsymbol{S}_G\to\boldsymbol{P}_S^{\mathrm{T}}\boldsymbol{S}_G,\quad \boldsymbol{E}_G\to\boldsymbol{P}_E^{\mathrm{T}}\boldsymbol{E}_G,$$

where $\boldsymbol{P}_S\in\mathbb{R}^{d_S\times d}$ and $\boldsymbol{P}_E\in\mathbb{R}^{d_E\times d}$ are the projection matrices, $\boldsymbol{P}_S^{\mathrm{T}}\boldsymbol{P}_S=\boldsymbol{P}_E^{\mathrm{T}}\boldsymbol{P}_E=\boldsymbol{I}_d$, and d is the subspace dimension. With the development of convex optimization methods, this function can be easily solved by off-the-shelf solvers. The reconstruction error term can be reconstructed with projection matrices as

$$\Phi_R(\boldsymbol{D},\boldsymbol{X}_S,\boldsymbol{X}_E)=\left\|\boldsymbol{P}_S^{\mathrm{T}}\boldsymbol{S}_G-\boldsymbol{D}\boldsymbol{X}_S\right\|_F^2+\left\|\boldsymbol{P}_E^{\mathrm{T}}\boldsymbol{E}_G-\boldsymbol{D}\boldsymbol{X}_E\right\|_F^2, \tag{3}$$

* Corresponding author (email: steven.xpw@ncu.edu.cn, a.g.song@seu.edu.cn)

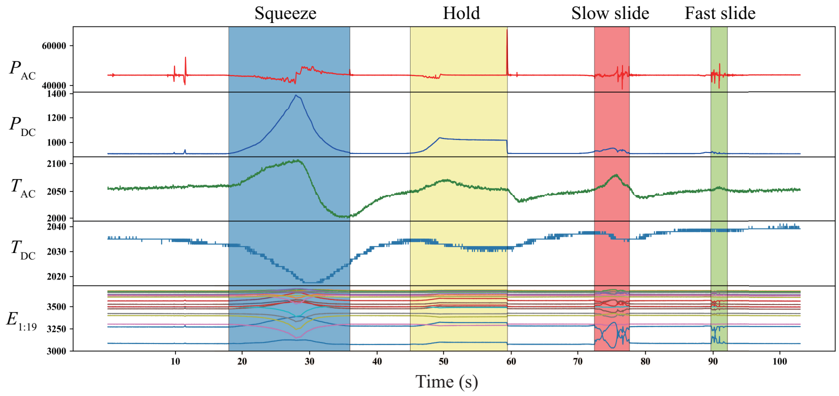

Figure 1 (Color online) Haptic signals were recorded from a PR2 robot trial using two BioTac sensors installed in its left hand. A single record contains pressure ($P_{\rm AC}$, $P_{\rm DC}$), temperature ($T_{\rm AC}$, $T_{\rm DC}$), and spatially distributed impedance-measuring electrodes ($E_{1:19}$), and each integrated process contains four exploratory procedures, including squeeze, hold, slow slide, and fast slide.

where $\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_K] \in \mathbb{R}^{d\times K}$ is a dictionary matrix with $d < K$. The proposed fusion model can be defined as the following optimization problem:

$$\min_{\boldsymbol{D},\boldsymbol{X}_S,\boldsymbol{X}_E,\boldsymbol{P}_S,\boldsymbol{P}_E} \Phi_R\left(\boldsymbol{D},\boldsymbol{X}_S,\boldsymbol{X}_E\right) + \Phi_P\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right). \quad (4)$$

$\|\boldsymbol{d}_k\|_2 \leqslant 1$ is a typical constraint that can prevent obtaining an oversized solution, and $\boldsymbol{P}_S^{\rm T}\boldsymbol{P}_S = \boldsymbol{P}_E^{\rm T}\boldsymbol{P}_E = \boldsymbol{I}_d$.

For our haptic adjective recognition model, the constraint of the proposed fusion model can be defined as

$$\Phi_P\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right) = \partial\left\|[\boldsymbol{X}_S,\boldsymbol{X}_E]\right\|_{1,1} + \Phi_L\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right), \quad (5)$$

where $\|[\boldsymbol{X}_S,\boldsymbol{X}_E]\|_{1,1}$ is the sum of absolute values of elements in $[\boldsymbol{X}_S,\boldsymbol{X}_E]$. The label of training samples is defined as

$$q_{*,i} = \begin{cases} +\delta, & \text{when object } i \text{ is labeled as positive,} \\ -\delta, & \text{when object } i \text{ is labeled as negative,} \end{cases} \quad (6)$$

where δ is a positive number usually set as one, and $*$ refers to any sample in object i. Labels of training samples can be combined as

$$\boldsymbol{Q} = \left[q_{1,1}, q_{2,1}, \ldots, q_{N_1,1}, \ldots, q_{N_2,2}, \ldots, q_{N_o,O}\right]. \quad (7)$$

Inspired by [6], we linearly transform all coefficient vectors to a common label for pairing, which can be defined as

$$\begin{aligned} +\delta &= \boldsymbol{W}_{S,T}\boldsymbol{x}_{S,T} = \boldsymbol{W}_E^{\rm T}\boldsymbol{x}_{E,T}, \\ -\delta &= \boldsymbol{W}_S^{\rm T}\boldsymbol{x}_{S,F} = \boldsymbol{W}_E^{\rm T}\boldsymbol{x}_{E,F}, \end{aligned} \quad (8)$$

where $\boldsymbol{x}_{S,T}$ is the coefficient vector of scalar signals of any positively labeled sample, and $\boldsymbol{x}_{S,F}$ is the coefficient vector of scalar signals of any negatively labeled sample. Similarly, $\boldsymbol{x}_{E,T}$ is the coefficient vector of electrode array signals of any positively labeled sample, and $\boldsymbol{x}_{E,F}$ is the coefficient vector of electrode array signals of any negatively labeled sample. According to the above discussion, the entire optimization problem can be constructed as

$$\begin{aligned} &\min_{\boldsymbol{D},\boldsymbol{X}_S,\boldsymbol{X}_E,\boldsymbol{P}_S,\boldsymbol{P}_E} \Phi_R\left(\boldsymbol{D},\boldsymbol{X}_S,\boldsymbol{X}_E\right) + \Phi_P\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right), \\ &\Phi_R\left(\boldsymbol{D},\boldsymbol{X}_S,\boldsymbol{X}_E\right) = \left\|\boldsymbol{P}_S^{\rm T}\boldsymbol{S}_G - \boldsymbol{D}\boldsymbol{X}_S\right\|_F^2 \\ &\qquad + \left\|\boldsymbol{P}_E^{\rm T}\boldsymbol{E}_G - \boldsymbol{D}\boldsymbol{X}_E\right\|_F^2, \\ &\Phi_P\left(\boldsymbol{X}_S,\boldsymbol{X}_E\right) = \partial\left\|[\boldsymbol{X}_S,\boldsymbol{X}_E]\right\|_{1,1} \\ &\qquad + \beta\left\|\begin{bmatrix}\boldsymbol{Q}\\ \boldsymbol{Q}\end{bmatrix} - \begin{bmatrix}\boldsymbol{W}_S^{\rm T}\boldsymbol{X}_S\\ \boldsymbol{W}_E^{\rm T}\boldsymbol{X}_E\end{bmatrix}\right\|_F^2 + \chi\left\|[\boldsymbol{W}_S,\boldsymbol{W}_E]\right\|_F^2, \end{aligned} \quad (9)$$

where $\|[W_S, W_E]\|_F^2$ can prevent the overfitting problem. β and χ are weight parameters. The above approximation solution can be obtained by iterating a process that immobilizes different variables according to a fixed sequence. With the optimal solution, $\boldsymbol{D}^*$, $\boldsymbol{W}_S^*$, $\boldsymbol{W}_E^*$, $\boldsymbol{P}_S^*$, and $\boldsymbol{P}_E^*$ are obtained. With the feature $\boldsymbol{s}_G$ of the scalar signal and the feature $\boldsymbol{e}_G$ of the electrode array signal of a testing sample, the label l of the testing sample can be decided by

$$l = \begin{cases} \text{positive}, & \boldsymbol{W}_S^{*\rm T}\boldsymbol{x}_S^* + \boldsymbol{W}_E^{*\rm T}\boldsymbol{x}_E^* \geqslant 0, \\ \text{negative}, & \boldsymbol{W}_S^{*\rm T}\boldsymbol{x}_S^* + \boldsymbol{W}_E^{*\rm T}\boldsymbol{x}_E^* < 0. \end{cases} \quad (10)$$

Conclusion. Traditional methods cannot fuse multimodal haptic measurements well; to solve this problem, we attempt to translate multimodal sparse codes obtained using a unified dictionary into a shared label. The proposed model not only preserves the original multimodal information but also transforms the raw data into a shared feature space.

Acknowledgements This work was partially supported by National Natural Science Foundation of China (Grant Nos. 62163024, 61903175, 61663027) and Academic and Technical Leaders Foundation of Major Disciplines of Jiangxi Province (Grant No. 20204BCJ23006).

References

1 Liu H, Yu Y, Sun F, et al. Visual-tactile fusion for object recognition. IEEE Trans Automat Sci Eng, 2017, 14: 996–1008

2 Strese M, Schuwerk C, Iepure A, et al. Multimodal feature-based surface material classification. IEEE Trans Haptics, 2017, 10: 226–239

3 Gao Y, Hendricks L A, Kuchenbecker K J, et al. Deep learning for tactile understanding from visual and haptic data. In: Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA), 2016. 536–543

4 Chu V, McMahon I, Riano L, et al. Robotic learning of haptic adjectives through physical interaction. Robot Auton Syst, 2015, 63: 279–292

5 Huang D, Wang Y F. Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In: Proceedings of 2013 IEEE International Conference on Computer Vision, 2013. 2496–2503

6 Jiang Z, Lin Z, Davis L S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. 1697–1704