Haptic Object Recognition: A Recurrent Approach

Chang Liu[1], Fuchun Sun[2] and Alan Yuille[3]

Abstract

The ability to recognize objects from haptic signal is vital for robots to interact with its environment, especially when other perception readings, like vision, are not available or not reliable. For example, when the robot performs a fine grasp task, the vision may likely fail to capture the small parts of the objects due to scale, occlusion or lighting. The sensor readings from haptic alone yet make it very hard for the robot to recognize the objects correctly, but considering the fact that human is capable of recognizing objects by just touching and holding objects, it is still desirable and interesting to investigate whether robots can achieve such ability.

Nonetheless, the task of haptic object recognition is not easy, and the difficulties lie in the sequential and high-dimensional characteristics of the haptic data. Compared to static data like images, haptic data incorporate time dimension, which means the consecutive frames of the captured data from the haptic sensors are highly correlated over time.

Deep learning has shown its effectiveness in numerous research areas, such as image classification in vision domain, speech recognition in auditory domain, machine translation in language domain. Inspired by the success of these research ares, in the literature of haptic domain, some efforts also have been made to explore the capabilities of deep learning techniques. Lele et al.[2] proposed a randomized tiling convolutional networks for haptic object recognition, where decomposed haptic data into spatial and temporal threads. Yang et al.[3] leveraged a four-layer convolutional neural network for haptic data representation. Haitian et al.[4] utilized fully convolutional network for material surface classification.

Despite the good performance in these works, it is worth noting that the ability of convolutional neural network attributes to its representation of local correlated information, thus makes it more suitable for vision related tasks such as image classification and object detection. When it comes to time-related sequential data, however, convolutional neural networks may find hard to handle, especially if data is of variable length. On the other hand, recurrent neural networks is a natural deep learning architecture for dealing with sequential data. Inspired by our previous work [5], we propose a recurrent neural networks based model for haptic object recognition. Similar idea exists in previous works, e.g., [1] and [2], yet our proposed model is distinct as (1) we propose an end-to-end trainable recurrent neural network, without the incorporation of deep Q network as in [1], or convolutional neural networks as in [2], and (2) we introduce an attention mechanism for the sequential data, aiming at learning the different contributions of each frame of the haptic data to the final categorization result.

Specifically, the proposed model is a supervised learning based model. The input to the model is the sequential readings from haptic senors of the robot hand, and each sequence is given a label. The model is mainly composed of recurrent neural networks, where on each time step, the network takes in the flattened readings from all the sensors. In consideration of the gradient vanishing problem, we choose LSTM unit as the core of the recurrent neural networks[6]. To cope with the long-term dependence of contextual information, we introduce an attention mechanism which learns to distribute attention to different hidden states of the recurrent neural networks. The intuition behind the attention mechanism is, not all the haptic readings are meaningful to current object categorization, in other words, the frames of the sensor readings do not contribute to the final result equally. Thus it is desirable for the model to distinguish the differences between each frames, by learning the attention weights that determine the contribution of each input frame. The proposed method is validated on popular haptic object recognition datasets, i.e., SD5, SD10, SPr7, SPr10, BDH5, BDH10 and HCs10, and achieves the state-of-the-art performance.

Acknowledgment

This work is jointly supported by National Natural Science Foundation of China under with Grant No.61621136008, 61327809,61210013,91420302 & 91520201. This paper is also supported by NSF STC award CCF-1231216.

References

[1] Sung, Jaeyong, J. Kenneth Salisbury, and Ashutosh Saxena. "Learning to Represent Haptic Feedback for Partially-Observable Tasks", ICRA, 2017

[2] Lele Cao, Ramamohanarao Kotagiri, et al. "Efficient spatio-temporal tactile object recognition with randomized tiling convolutional networks in a hierarchical fusion strategy", AAAI, 2016

[3] Yang Gao, Lisa Hendricks, et al. "Deep learning for tactile understanding from visual and haptic data", ICRA, 2016

[4] Haitian Zheng, Lu Fang, et al. "Deep learning for surface material classification using haptic and visual information", IEEE Transactions on Multimedia, 2016

[5] Chang Liu, Fuchun Sun, et al. "MAT: A Multimodal Attentive Translator for Image Captioning", IJCAI, 2017

[6] Sepp Hochreiter and Jurgen Schmidhuber, "Long short-term memory", Neural computation, 1997

[1]Chang Liu is with the Department of Computer Science, Tsinghua University, and the Department of Cognitive Science and Computer Science, Johns Hopkins University, cliu13@mails.tsinghua.edu.cn

[2]Fuchun Sun is with the Department of Computer Science, Tsinghua University, fcsun@mail.tsinghua.edu.cn

[3]Alan Yuille is with the Department of Cognitive Science and Computer Science, Johns Hopkins University, alan.yuille@jhu.edu