Multi-Modal Learning for Dynamic Tactile Sensing

Oliver Kroemer1, Christoph H. Lampert2, and Jan Peters1
1 Technische Universitaet Darmstadt, Germany. {kroemer,peters}@ias.tu-darmstadt.de
2 Institute of Science and Technology Austria, Austria. chl@ist.ac.at

I. INTRODUCTION

Dynamic tactile sensing allows humans to infer surface and material properties from the vibrations caused by the sliding motion between the skin and an object [1]. For example, one can easily determine the roughness of a surface by sliding one’s finger tip over the surface [2]. This sensory modality is also fundamental for tool usage, as it detects the vibrations resulting from the tool making contact with another object [3], and senses incipient slip between the fingers and the tool in order to maintain a good grip [4]. Robots performing dexterous manipulation tasks could therefore gain similar benefits from using dynamic tactile sensing [5], [6], [7].

However, raw time-series data received from dynamic tactile sensors is usually noisy and high-dimensional. The signal will also often contain confounding vibrations from other sources, e.g., the robot’s own vibrations [6]. Hence, it is difficult to directly use this sensory information to discriminate between different surfaces. Instead, we propose first learning a low-dimensional representation of the data. This low-dimensional space should capture the vibrations caused by the textured surfaces, but exclude dimensions containing the additional vibrations that are irrelevant to the tactile sensing task.

In order to determine which components of the signal correspond to the textured surface, we employ a human-inspired approach. Humans are capable of combining information from multiple sensor modalities, such as touch, vision, and audition [8], [3]. Similarly, robots can learn a suitable low-dimensional representation for dynamic tactile data, by using vision information. In particular, the robot can determine relevant features by finding correlations between the vision and the tactile data of various textured surfaces. Parts of the data relating to the texture of the surface will occur in both sensor modalities. However, the additional vibrations detected by the dynamic tactile sensor will not be captured by the images of the objects, and hence these components will be excluded.

In this work, we present two algorithms for learning low-dimensional representations of dynamic tactile data. These methods use the correlations between tactile and vision data taken from the same surfaces in order to determine which dimensions are relevant to the tactile sensing task. Once a lower-dimensional space has been learned, the robot can directly represent new tactile data in this space without requiring corresponding vision data. Using the proposed approach, the robot (see Fig. 1) was able to accurately discriminate between various textured surfaces using only the basic dynamic tactile sensor shown in Fig. 2.

II. MULTI-MODAL LEARNING

The robot should learn a low-dimensional representation of the tactile information by combining both tactile and vision data. However, acquiring vision data that is perfectly paired to corresponding tactile samples is not trivial, as the object will often shift between the tactile exploration and the visual inspection. Acquiring good pairings between samples is particularly difficult in unstructured environments. Therefore, we only assume that samples are weakly paired between the two sensor modalities; i.e. instead of specifying one-to-one pairings, sets of multiple tactile samples are paired with sets of multiple vision samples. For example, all of the tactile samples acquired from one object may be paired to all of the vision data taken from the same object.

The first of our two algorithms is called mean maximum covariance analysis ($\mu$MCA). The $\mu$MCA approach represents each set of weakly-paired samples my its mean. These mean
samples are then strongly paired between the modalities, and maximum covariance analysis is used to compute the low-dimensional space for representing the tactile data. The \( \mu \)MCA method can be easily updated when new data is obtained [9].

The second method is called weakly-paired maximum covariance analysis (WMCA), and attempts to automatically determine strong pairings between individual samples [9], [10]. This approach uses an alternating maximization scheme in order to maximize the covariance between the tactile and vision data. In the first step, the pairing of samples between modalities is treated as a linear assignment problem based on the covariances between the samples [11]. In the second step, a maximum covariance analysis is performed on the paired data in order to compute a lower dimensional space. The data is projected into this lower dimensional space, and the pairings between samples are recomputed. This process is repeated until the pairings remain unchanged and, hence, a suitable dimensionality reduction can be found. The WMCA algorithm is generally more robust to outliers than \( \mu \)MCA [9].

III. EXPERIMENTS

The proposed methods were tested using the robot shown in Fig. 1, which explored 17 different surface types during the experiments. The dynamic tactile sensor, shown in Fig. 2, consists of a pin and a microphone. By sliding the pin across textured surfaces, the robot can generate vibrations in the pin that the microphone can detect and record. Example data collected by the robot for two types of surfaces can be seen in Fig. 3.

The goal of the robot was to learn to classify the different types of textured surfaces. The learning was performed in two phases. In the first phase, the robot was allowed to learn a lower-dimensionalsional representation using both vision and tactile information. In the second phase, the robot had to use the learned dimensionality reduction to classify different surfaces without using vision data. For comparison, the classification task was also performed using PCA to directly compute lower-dimensional representations from only the tactile data.

The results of the classification experiment are shown in Fig. 4. Both \( \mu \)MCA and WMCA allowed the robot to accurately classify the textured surfaces. The results show that the use of multi-modal learning can significantly improve the robot’s ability to distinguish between different types of surfaces.

ACKNOWLEDGMENTS

The project receives funding from the European Community’s Seventh Framework Programme under grant agreement no. ICT- 248273 GeRT.

REFERENCES