Unsupervised Learning of Spatio-Temporal Features for Tactile Data

Marianna Madry Liefeng Bo Danica Kragic Dieter Fox

Introduction

Tactile feedback obtained during grasping of an object is crucial for a number of tasks in robotics. It allows to identify object properties and recognize its class, or to localize an object in a robot hand and assess quality and stability of the applied grasp [1]–[3]. The accomplishment of these tasks greatly depends on the choice of data representation, i.e. features that are used to capture signal properties, as it was was shown for visual data [4]–[7]. The representation of tactile data should not only provide robustness to real-world conditions, but also serve a wide variety of applications and be flexible to adapt to their specific requirements.

In previous works tactile signal have been typically represented using a manually crafted set of features based on prior knowledge about the properties of the inputs. Early works aimed at identifying simple primitive shapes in object imprints in the tactile matrices (points and lines [8], [9]). Recent works use higher level geometric properties of pressure patterns, such as their position, area or higher-order moments [1], [3], [10], [11]. Other approaches build on features developed at Computer Vision, such as SIFT [12]. A very limited amount of work have aimed at unsupervised extraction of features by applying to tactile data the K-means algorithm or covariance analysis [13]–[15].

Spatio-Temporal Hierarchical Matching Pursuit

Using of a *pre-defined* set of feature can limit capability of the representation to serve different applications and may results in capturing characteristics that have minor importance for a task [11]. In contrast, in this work we generate representation of tactile data directly from raw input signals without the need of specifying discriminative characteristics a-priori. This may be achieved using *unsupervised feature learning techniques* and *deep learning methods* [5]. We propose the *Spatio-Temporal Hierarchical Matching Pursuit (ST-HMP)* that in an unsupervised manner learns properties of a *time* series of tactile sensor measurements, see Fig. 1.

The ST-HMP is based on the concept of hierarchical feature learning realized using sparse coding. It builds on the recently introduced the Hierarchical Matching Pursuit (HMP) [16] that is a multilayer sparse coding network

M. Madry and D. Kragic are with the Center for Autonomous Systems and the Computer Vision & Active Perception Lab, CSC, KTH Royal Institute of Technology, Stockholm, Sweden. {madry,danik}@csc.kth.se Liefeng Bo is with Amazon Inc. The work was completed while he was affiliated with the Intel Science and Technology Center on Pervasive Computing, Seattle, USA. {liefengbo@gmail.com}. Dieter Fox is with the Department of Computer Science & Engineering, University of Washington, Seattle, USA. {foxg@cs.washington.edu}.

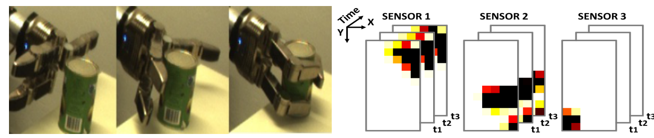

Fig. 1. (*Left*) Example of grasp execution with the three finger Schunk Dextrous Hand (SDH) [3]. (*Right*) Tactile measurements create a sequence over time in which information in consecutive frames is correlated.

creating a representation from raw data, layer by layer, with an increasing receptive field size. This approach has been successfully used for visual object recognition [17] and as such is limited to spatial signals (images). However, in grasping temporal information is crucial for good performance [2]. We propose a method to extend the HMP framework to *spatio-temporal* feature learning.

In order to compute the ST-HMP, we first sample a large collection of spatial (or spatio-temporal) patches from tactile matrices and use them to learn codebooks in an unsupervised fashion by applying K-SVD With the learned codebooks, the ST-HMP computes sparse code features using the orthogonal matching pursuit (OMP). Then, features extracted from consecutive frames are repeatedly max pooled over the space and time dimensions at several scales with an increasing size of a receptive field (cell) creating a spatio-temporal pyramid, see Fig. 2. A final feature vector for the whole tactile sequence is the concatenation of aggregated sparse codes in each spatio-temporal cell. For algorithm details, we refer the reader to our recent publication [7].

Experimental Evaluation

In order to demonstrate universal properties of the ST-HMP we considered two typical classification problems for tactile data: (1) grasp stability assessment and (2) object instance recognition. We performed extensive evaluation on six synthetic and real databases that were collected using the Schunk Dexterous, Schunk Parallel and iCub hands for several sets of objects of widely diverse characteristics. The robot hands and objects sets are presented in Fig. 3.

First, we evaluated the effect of encoding not only spatial, but also temporal information. Thus, we separately applied the original spatial HMP to each frame in a grasping sequence, and then added temporal information by computing the ST-HMP descriptor. The results confirmed that adding temporal dimension highly improved accuracy for the HMP for all considered cases, see Fig 4.

Second, we compared the ST-HMP with the state-of-the-art methods that in different ways encode temporal information, such as Hidden Markov Models (HMMs) [3] [18], Gaussian Processes (GP) with recursive kernels [2], Dynamic

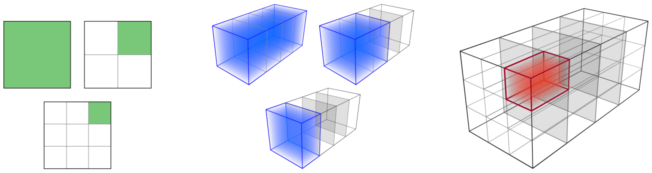

Fig. 2. Schematic illustration of a partition of data in: (*Left*) the spatial pyramid with the partition 1-2-3, (*Middle*) the temporal pyramid with the partition 1-2-4, and (*Right*) the smallest possible cell in the given spatio-temporal pyramid. The size of a cell in which features are max-pooled at each pyramid level is marked with a color. In the 3-level spatial pyramid the matrix is divided into $S = 1 + 2^2 + 3^2$ cells giving the original HMP of the size $14 \times M$, where M is the size of the codebook. The 3-level temporal pyramid partition gives $T = 1 + 2 + 2^2$ cells. In this case, the dimensionality of the ST-HMP feature vector obtained for the whole tactile sequence for the spatio-temporal pyramid pooling is equal to $14 \times 7 \times M$.

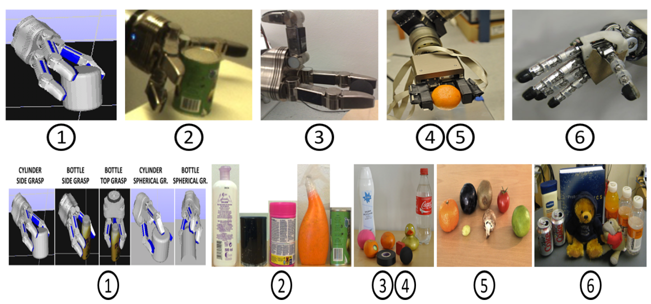

Fig. 3. (*Top*) Five robot hands used to collect databases. From left: the model of the 3-finger Schunk Dexterous Hand (SDH), 3-finger SDH, 3-finger SDH with Weiss tactile sensors, 2-finger Schunk Parallel Gripper, 5-finger iCub hand. (*Bottom*) Object sets used to collect databases. The numbers below the images indicate the database collected using the hand and object set depicted above. All databases are annotated with an object class and the databases (1) and (2) additionally with information about grasp stability. Figures are reproduced with permission from [1], [2], [18].

Time Warping (DTW) [1], decision trees [2], and techniques based on SVMs and AdaBoost [18] [2]. In all these cases *pre-defined* types of features have been used, such as geometric properties of contact regions or higher order moments. The ST-HMP outperformed the previously published methods by a large margin, see Fig 4. Please note the variety of the analyzed cases, i.e. different classification tasks and types of data collected for different object sets and robot hands. The detailed analysis can be found in [7].

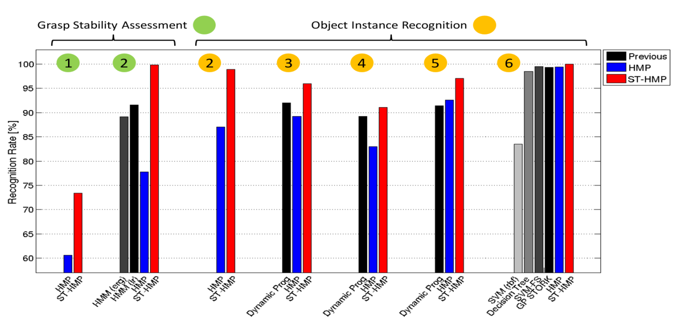

Fig. 4. Experimental comparison of the proposed ST-HMP with the original HMP, and the previously published results [1]–[3], [18]. The numbers above the images indicate one of the six database id. We evaluated our method for two tasks: grasp stability assessment (1-2 bar groups) and object instance recognition (3-7 bar groups). Support Vector Machines (SVMs) with a linear kernel were used as a classier in all cases. We closely followed the setups from the previous publications.

Conclusions and Future Work

We proposed a new descriptor named Spatio-Temporal Hierarchical Matching Pursuit (ST-HMP) that captures properties of a time series of tactile measurements. It is based on the concept of unsupervised hierarchical feature learning and extracts rich spatio-temporal structures from raw tactile data without the need to predefine discriminative data characteristics. An extensive evaluation on several synthetic and real databases showed that the ST-HMP can be successfully applied to tactile data originating from different robot hands and objects, and is an universal descriptor that can be applied to different classification tasks. In the future, we plan to investigate a principled way to make the ST-HMP applicable to multi-modal data, such as finger positions and joint angles.

Acknowledgements

This work was supported by the Swedish Foundation for Strategic Research, the EU FP7 project RoboHow.Cog (FP7-ICT-288533) and the Intel Science and Technology Center for Pervasive Computing (ISTC-PC). We thank Yasemin Bekiroglu, Alin Drimus, Jimmy Jørgensen and Harold Soh for providing access to the databases.

References

[1] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic. Design of a flexible tactile sensor for classification of rigid and deformable objects. *RAS*, 2012.

[2] H. Soh, Y. Su, and Y. Demiris. Online spatio-temporal gaussian process experts with application to tactile classification. In *IROS*, 2012.

[3] Y. Bekiroglu, D. Kragic, and V. Kyrki. Learning grasp stability based on tactile data and hmms. In *RO-MAN*, 2010.

[4] S. Edelman. *Representation and recognition in vision*. MIT press, 1999.

[5] Y. Bengio, A. C. Courville, and P. Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR*, 2012.

[6] M. Madry, C. H. Ek, R. Detry, Kaiyu Hang, and D. Kragic. Improving generalization for 3D object categorization with Global Structure Histograms. In *IROS*, 2012.

[7] M. Madry, L. Bo, D. Kragic, and D. Fox. ST-HMP: Unsupervised Spatio-Temporal Feature Learning for Tactile Data. In *ICRA*, 2014.

[8] M. Okamura and M. R Cutkosky. Haptic exploration of fine surface features. In *ICRA*, 1999.

[9] R. A. Russell. Object recognition by a smart tactile sensor. In *ACRA*, 2000.

[10] N. Gorges, S. E. Navarro, D. Goger, and H. Worn. Haptic object recognition using passive joints and haptic key features. In *ICRA*, 2010.

[11] M. Schopfer, M. Pardowitz, R. Haschke, and H. Ritter. Identifying relevant tactile features for object identification. *STAR*, 2012.

[12] Z. Pezzementi, E. Plaku, C. Reyda, and G.D. Hager. Tactile-object recognition from appearance information. *TOR*, 2011.

[13] Z. Pezzementi, C. Reyda, and G. D. Hager. Object mapping, recognition, and localization from tactile geometry. In *ICRA*, 2011.

[14] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard. Object identification with tactile sensors using bag-of-features. In *IROS*, 2009.

[15] Hongbin Liu, Xiaojing Song, T. Nanayakkara, L. D. Seneviratne, and K. Althoefer. A computationally fast algorithm for local contact shape and pose classification using a tactile array sensor. In *ICRA*, 2012.

[16] L. Bo, X. Ren, and D. Fox. Hierarchical Matching Pursuit for Image Classification: Architecture and Fast Algorithms. In *NIPS*, 2011.

[17] L. Bo, X. Ren, and D. Fox. Multipath sparse coding using hierarchical matching pursuit. In *CVPR*, 2013.

[18] Y. Bekiroglu, J. Laaksonen, J. A. Jorgensen, V. Kyrki, and D. Kragic. Assessing grasp stability based on learning and haptic data. *TROB*, 2011.