UNSUPERVISED SPATIO-TEMPORAL FEATURE LEARNING FOR TACTILE DATA

Marianna Madry[1] Liefeng Bo[2,3] Danica Kragic[1] Dieter Fox[3]

[1] CVAP Lab, KTH – Royal Institute of Technology, Stockholm, Sweden

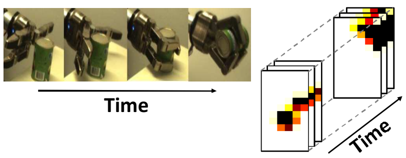
[2] ISTC-Pervasive Computing Intel Labs, Seattle, USA

[3] RSE Lab, University of Washington, Seattle, USA

KEYWORDS: TACTILE, UNSUPERVISED FEATURE LEARNING, DEEP LEARNING, TEMPORAL SIGNAL, GRASP STABILITY, OBJECT RECOGNITION

Series of Tactile Data

- create a sequence of matrices over time
- reflects object and grasp properties
- is useful for a number of applications



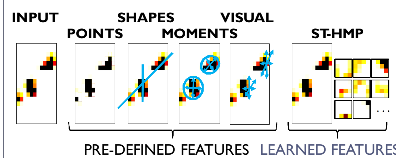
Flexible Representation

How to design a general and flexiable representation that is useful for various applications and data types?

Contribution

A new descriptor ST-HMP for unsupervised feature learning for time series of tactile measurements

ST-HMP vs. Previous Work



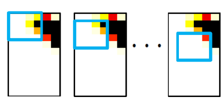
- Pre-defined: can be irrelevant, inflexible
- ST-HMP: learns from raw input, no need to specify characteristics a priori

Spatio-Temporal Hierarchical Matching Pursuit (ST-HMP)

STEP 1: SIGNAL REPRESENTATION

- Sampling patches using a sliding window

OBSERVED SIGNAL Y



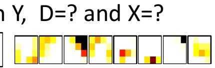
- ST-HMP represents a signal as sparse linear combination of codewords

Y (OBSERVED SIGNAL) $=$ D (OVERCOMPLETE DICTIONARY) X (SPARSE CODES)

STEP 2: CODEBOOK LEARNING

- Problem: given Y, D=? and X=?

DICTIONARY D

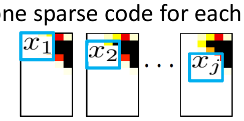


SPARSE CODES X $[x_1, \cdots, x_n, \cdots, x_N]$

- Underdetermined system of equations is solved by minimizing the error using K-SVD

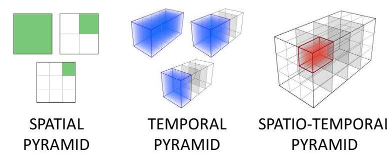
$$\min_{D,X} \|Y - DX\|_F^2$$

- Result: one sparse code for each patch



STEP 3: CODE AGGREGATION

- Sparse codes are max-pooled in space & time in cells of an increasing size

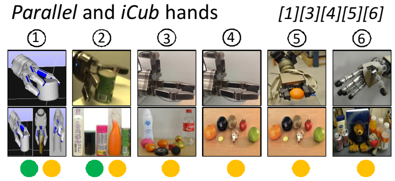


- Feature vector = concatenated aggregation results for each cell

Experimental Evaluation

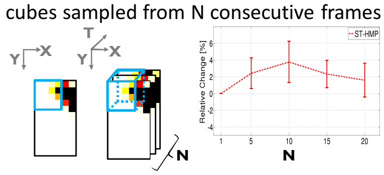
Databases

- Applications: grasp stability assessment; object recognition
- Six databases: *Schunk Dexterous*, *Schunk Parallel* and *iCub* hands [1][3][4][5][6]



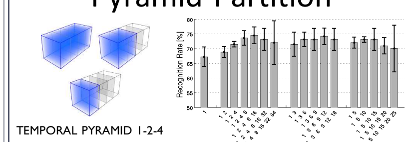
ST-Dictionary

- Dictionary learned for 3D spatio-temporal cubes sampled from N consecutive frames



- Improves until information captured by the ST-dic and ST-pyramid is redundant

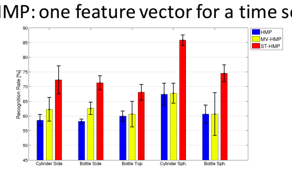
Pyramid Partition



- Using multiple levels significantly improves recognition rate
- Allows to adapt to processes of different temporal resolution

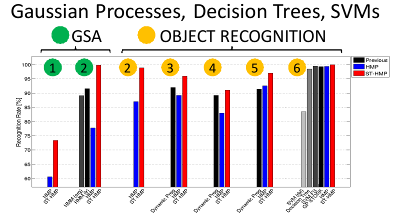
Encoding ST Information

- HMP: computed for every frame (only spatial)
- MV-HMP: majority voting over HMP results (temporal information weakly captured)
- ST-HMP: one feature vector for a time series



ST-HMP vs. Previous Work

- Comparison with HMMs, Dynamic Prog. Gaussian Processes, Decision Trees, SVMs



Conclusions

Extensive evaluation showed that ST-HMP:

- **outperforms by a large margin**
  - methods ignoring temporal component
  - state of the art: HMMs, DP, GP, DT, SVMs
- **is a universal descriptor** that can be successfully applied to:
  - different applications
  - data collected using different robot hands and objects

[1] Madry et al. "ST-HMP: Unsupervised Spatio Temporal Feature Learning for Tactile Data". In ICRA, 2014.

[2] Bo et al. "Learning Hierarchical Sparse Features for RGB-(D) Object Recognition". In IJRR, 2014

[3] Bekiroglu et al. "Assessing grasp stability based on learning and haptic data". In TORB, 2011.

[4] Soh et al. "Online spatio-temporal GP experts with application to tactile classification.". In IROS'12

[5] Drimus et al. "Design of a flexible tactile sensor for classification of rigid and deformable objects.". RAS'12

[6] Bekiroglu et al. " Learning grasp stability based on tactile data and HMMs.". In ROMAN, 2010.