Online Activity Recognition through Kernel Methods

Evgeni Pavlidis

Advisor: Dr. Rudolph Triebel
Supervisor: Prof. Dr. Daniel Cremers

2014-12-05

Outline

Table of Contents

The Spencer Project

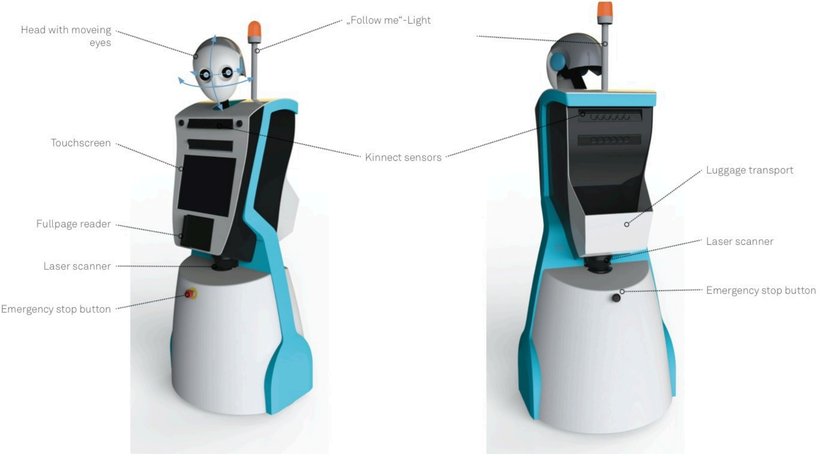


Figure: The final design of the SPENCER robot and its sensors.

Problem statement

Devise an algorithm which:

- can classify activities by pose sequences
- is capable of online recognition
- does not need large amount of training data
- can be integrated easily into ROS

Problem statement

Devise an algorithm which:

- can classify activities by pose sequences
- is capable of online recognition
- does not need large amount of training data
- can be integrated easily into ROS

Difficulties

Pose space is high dimensional

How to compare poses?

Classification of time-series data

How to classify sequences?

Benchmark

The Cornell Daily Living Activities Dataset (CAD-60) [6] is used for evaluation

- Consists of 4 persons (one left handed)
- Each person performs 12 activities several times
- The activities are: rinsing mouth, brushing teeth, wearing contact lens, talking on the phone, drinking water, opening pill container, cooking (chopping), cooking (stirring), talking on couch, relaxing on couch, writing on whiteboard, working on computer
- The data consists of:

RGB images, depth images, poses (as extracted by the OpenNI framework)

Gaussian Process - Latent Variable Model [1]

- A model for non-linear dimensionality reduction
- Defines a mapping from latent space to observed space
- Gives us a notion of certainty for the mapping

[1] Neil D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data.
In *Nips*, volume 2, page 5, 2003

Dual Probabilistic Principal Components Analysis

$$\mathbf{y}_i = W\mathbf{x}_i + n_i$$
$$n_i \sim \mathcal{N}(0, \sigma^2 I)$$
$$p(Y|X, W) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{y}_i | W\mathbf{x}_i, \sigma^2 I)$$

Probabilistic PCA

$$p(X) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{x}_i | 0, I)$$

With these priors we can integrate out the latent points:

$$p(Y|W) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{y}_i | 0, WW^T + \sigma^2 I)$$

Dual Probabilistic Principal Components Analysis

$$\boldsymbol{y}_i = W\boldsymbol{x}_i + n_i$$
$$n_i \sim \mathcal{N}(0, \sigma^2 I)$$
$$p(Y|X, W) = \prod_{i=1}^{n} \mathcal{N}(\boldsymbol{y}_i | W\boldsymbol{x}_i, \sigma^2 I)$$

Probabilistic PCA

$$p(X) = \prod_{i=1}^{n} \mathcal{N}(\boldsymbol{x}_i | 0, I)$$

With these priors we can integrate out the latent points:

$$p(Y|W) = \prod_{i=1}^{n} \mathcal{N}(\boldsymbol{y}_i | 0, WW^T + \sigma^2 I)$$

Dual Probabilistic PCA

$$P(W) = \prod_{i=1}^{d} \mathcal{N}(\boldsymbol{w}_i | 0, \boldsymbol{I})$$

With these priors we can integrate out the parameters:

$$p(Y|X) = \prod_{i=1}^{d} \mathcal{N}(\boldsymbol{y}_i | 0, XX^T + \sigma^2 I)$$

Gaussian Process - Latent Variable Model

We can interpret: $XX^T + \sigma^2 I = K$. as a linear kernel, leading to:

$$P(Y|X) = \prod_{i=1}^{d} \mathcal{N}(\boldsymbol{y}_i|0, K)$$

Thus the Dual PPCA can be interpreted as product of GPs with a linear kernel.

Gaussian Process - Latent Variable Model

We can interpret: $XX^T + \sigma^2 I = K$. as a linear kernel, leading to:

$$P(Y|X) = \prod_{i=1}^{d} \mathcal{N}(\boldsymbol{y}_i|0, K)$$

Thus the Dual PPCA can be interpreted as product of GPs with a linear kernel.

The log-likelihood is:

$$log\ p(Y|X) \propto -\frac{d}{2}\log(|K|) - \frac{1}{2}tr(K^{-1}YY^T)$$

where the covariance K depends on the latent points X and the hyper-parameters θ.

Back Constrained GP-LVM

Back-constraints [2]

Instead of optimizing the latent space directly, optimize the parameters of a mapping instead:

$$x_{i,j} = g_j(\boldsymbol{y}_i, \gamma)$$

$$g_j(\boldsymbol{y}_n, A, l, \sigma) = \sum_{i=1}^{n} A_{j,i} k(\boldsymbol{y}_n, \boldsymbol{y}_i)$$

Advantages:

- Gives us an inverse mapping (observed to latent space)
- Local distances are preserved

[2] Neil D. Lawrence and Joaquin Quiñonero-Candela. Local distance preservation in the GP-LVM through back constraints.
In *Proceedings of the 23rd international conference on Machine learning*, pages 513–520. ACM, 2006

Table of Contents

Daily Living Activity Recognition

RGB-D Camera-based Daily Living Activity Recognition. [3]

Feature extraction

A feature (pose) consists of:

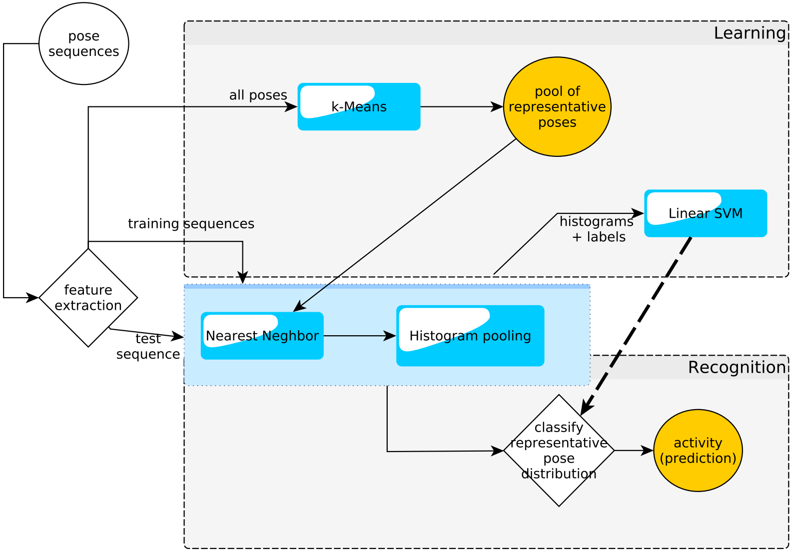
- a structural part (the differences between each joint pair)
- a motion part (the difference between current and previous frame for each joint)

Method

- Find the most representative poses from all data (k-Means algorithm)
- Quantize each sequence using these representative poses (nearest neighbor)
- Compute a distribution for each sequence (Bag-of-features)
- Learn a linear SVM on the distributions

[3] Chenyang Zhang and Yingli Tian. RGB-d camera-based daily living activity

Illustration



Extensions:

- Extract representative poses for each class
- Use sequence alignment functions for classification
  - Longest Common Subsequence
  - Dynamic Time Warping

Results

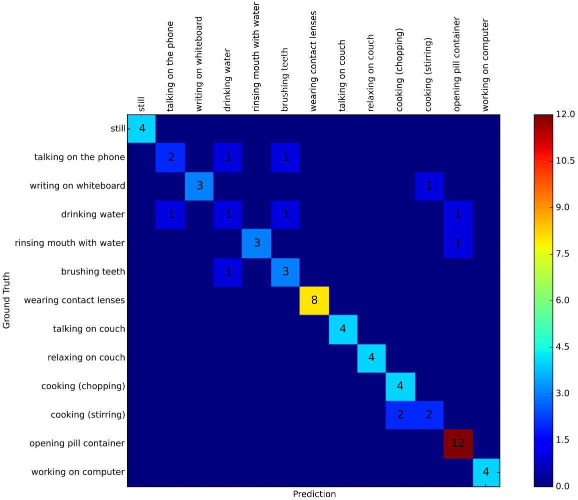


Figure: Bag-of-features approach with 128 clusters. precision 84%, recall 84%

k-Means with Longest Common Subsequence

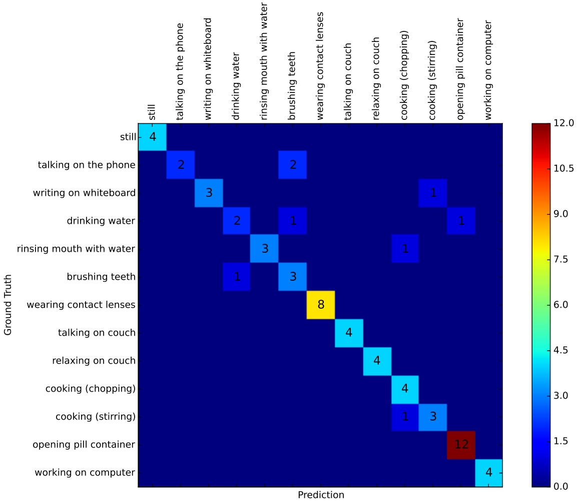


Figure: LCS approach with 64 clusters per class. precision 90%, recall 88%

Issues

Very difficult to adjust the algorithm to perform online recognition.

Discriminate Sequence Back-Constrained GP-LVM

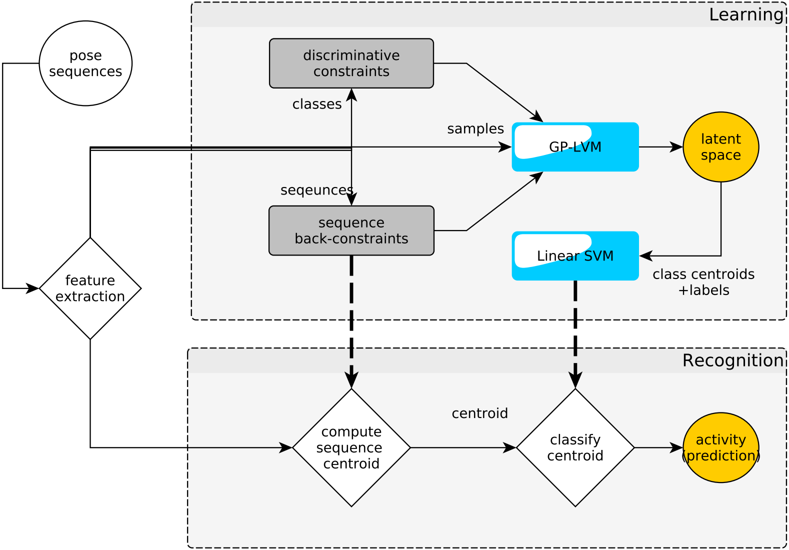
Discriminative Sequence Back-constrained GP-LVM [4]

- Perform GP-LVM based dimensionality reduction on all poses
- Recognition is done by classifying the centroid of a sequence in latent space (linear SVM)
- Make sure that similar sequences fall nearby in latent space (Sequence back-constraints)
- Learn a mapping which can map a sequence from the observed space to a centroid in the latent space (Sequence back-constraints)
- Make the clustering in the latent space more discriminative for the classes (Discriminative GP-LVM)

---

[4] Valsamis Ntouskos, Panagiotis Papadakis, and Fiora Pirri. Discriminative sequence back-constrained GP-LVM for MOCAP based action recognition:. pages 87–96. SciTePress - Science and and Technology Publications, 2013

Illustration



Issues

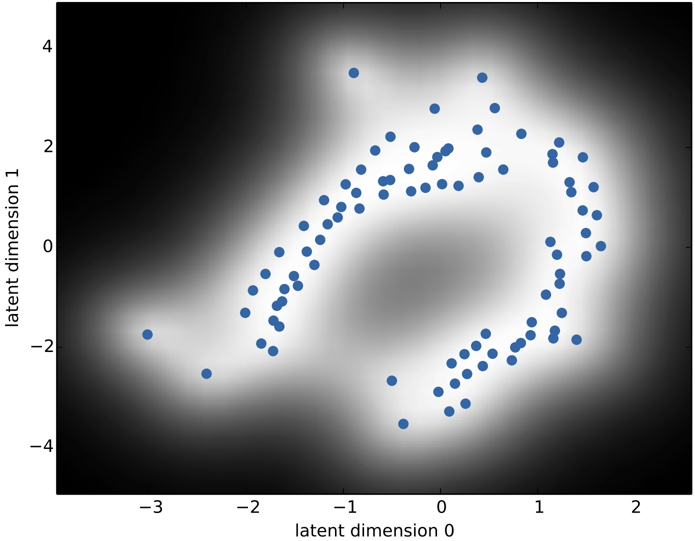
- Due to the nature of more complex activities and the huge search space, the optimization of the GP-LVM failed
- This can also be due to the DTW measure in the sequence alignment kernel, which does not represent an appropriate similarity measure for more complex activities

Inspired by: Gaussian Process Regression Flow [5]

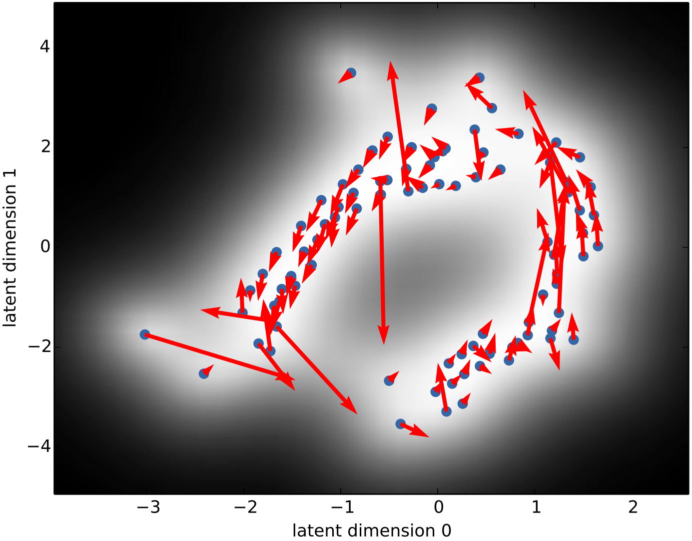
- Perform a separate dimensionality reduction for each activity class
- Learn a motion flow field by GP regression on the velocity function
- Online recognition by comparing the incoming motion with each flow field and updating the belief

[5] Kihwan Kim, Dongryeol Lee, and Irfan Essa. Gaussian process regression flow for analysis of motion trajectories.
In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1164–1171. IEEE, 2011

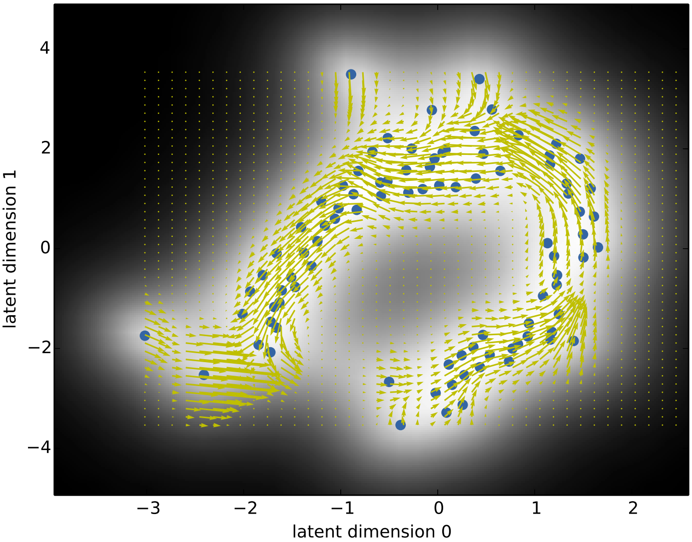
Latent Motion



Latent Motion - velocities



Latent Motion Flow Field



Illustration

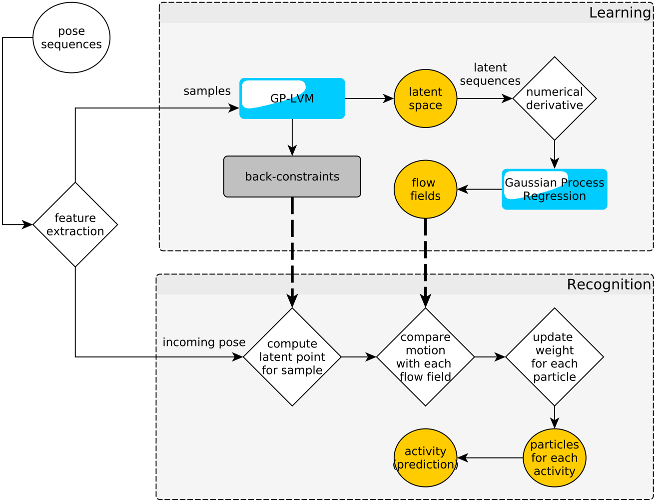


Figure: Illustration of the Gaussian Process - Latent Motion Flow approach.

Issues

- No smooth mapping from observed space to latent space
- This leads to discontinuities when learning the regression
- Possible solution:
    - Use spatio-temporal constraints in the optimization [6]

[6] Michał Lewandowski, Dimitrios Makris, and Jean-Christophe Nebel. Probabilistic feature extraction from multivariate time series using spatio-temporal constraints. In Joshua Zhexue Huang, Longbing Cao, and Jaideep Srivastava, editors, *Advances in Knowledge Discovery and Data Mining*, number 6635 in Lecture Notes in Computer Science, pages 173–184. Springer Berlin Heidelberg, January 2011

Table of Contents

Conclusions

- Local motion tendencies are more discriminative for complex activities then the overall dynamics
- Common dimensionality reduction for a large number of activities is extremely difficult
- Optimization of the *GP-LVM* is very difficult and strongly depends on the initialization

Contributions

- Implementation and extensions of an existing k-Means based approach in Python
- Implementation of a *ROS* module capable of activity recognition in real-time
- Implementation of a GUI client used to record, label and learn a model for new activities
- Implementation of the Discriminative Sequence Back-constraint GP-LVM in Python
- A novel approach for activity recognition using latent motion flow fields

Outlook

Implementation of the GP-Latent Motion Field using spatio-temporal GP-LVM

Use constraints based on Laplacian matrices for temporal and spatial graphs extracted from the time series

Semi-supervised activity learning by automatic segmentation of activities

Use the variances from the GP-LVM and the GP regression to identify unseen poses and motions

Thank you

Thank you for your attention!

BoF Approach with subsequences

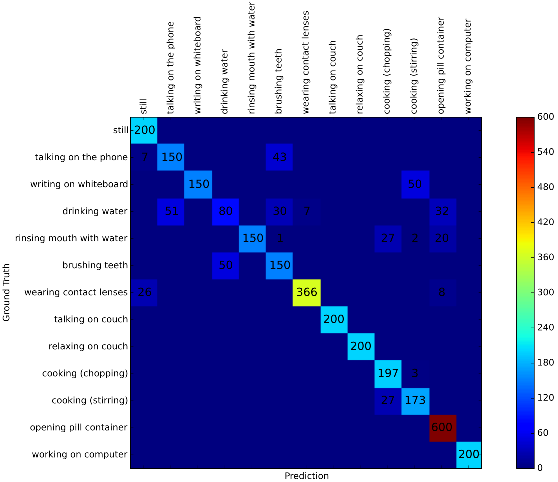


Figure: Confusion matrix: BoF approach with 128 clusters and intervals of 100 frames. Randomly sampled 50 times. precision 88%, recall 88%

Sequence Back-constraints

Define a similarity measure between sequences in observed space:

$$g_q(Y_s) = \sum_{m=1}^{S} a_{mq} k(Y_s, Y_m)$$

where the kernel is $k(Y_s, Y_m) = \gamma e^{\mathrm{DTW}(Y_s, Y_m)}$.
This measure is to be preserved in the latent spaces.

$$g_q(Y_s) = \mu_{sq} = \frac{1}{L_s} \sum_{n \in J_s} x_{nq}$$

Discriminative GP-LVM [7]

Make the latent space more discriminative by minimizing inner-class variance and maximizing inter-class separability.

- The distance between classes

$$S_b = \sum_{i=1}^{l} \frac{n_i}{n} (\boldsymbol{\mu_i} - \boldsymbol{\mu})(\boldsymbol{\mu_i} - \boldsymbol{\mu})^T$$

- The variance within the classes

$$S_w = \frac{1}{n} \sum_{i=1}^{l} \sum_{j=1}^{n_i} \frac{n_i}{n} (\boldsymbol{x_{i,j}} - \boldsymbol{\mu_i})(\boldsymbol{x_{i,j}} - \boldsymbol{\mu_i})^T$$

$$J(X) = tr(S_w^{-1} S_b)$$

[7] Raquel Urtasun and Trevor Darrell. Discriminative gaussian process latent variable model for classification.
In *Proceedings of the 24th international conference on Machine learning*, pages 927–934. ACM, 2007

References

Kihwan Kim, Dongryeol Lee, and Irfan Essa.
Gaussian process regression flow for analysis of motion trajectories.
In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1164–1171. IEEE, 2011.

Neil D. Lawrence.
Gaussian process latent variable models for visualisation of high dimensional data.
In *Nips*, volume 2, page 5, 2003.

Neil D. Lawrence and Joaquin Quiñonero-Candela.
Local distance preservation in the GP-LVM through back constraints.
In *Proceedings of the 23rd international conference on Machine learning*, pages 513–520. ACM, 2006.