Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Improving Robustness of Slow Feature Analysis Based Localization Using Loop Closure Events

Benjamin Metka, Ute Bauer-Wersing, Mathias Franzius

2016

Preprint:

This is an accepted article published in International Conference on Artificial Neural Networks (ICANN). The final authenticated version is available online at: https://doi.org/[DOI not available]

Improving Robustness of Slow Feature Analysis Based Localization Using Loop Closure Events

Benjamin Metka[1], Mathias Franzius[2], and Ute Bauer-Wersing[1]

[1] Frankfurt University of Applied Sciences,
Nibelungenplatz 1, 60318 Frankfurt am Main, Germany
bmetka@fb2.fra-uas.de
[2] Honda Research Institute Europe GmbH,
Carl-Legien-Strae 30, 63073 Offenbach, Germany

Abstract. Hierarchical Slow Feature Analysis (SFA) extracts a spatial representation of the environment by directly processing images from a training run and has been shown to enable self-localization of a mobile robot by encoding its position as slowly varying features. However, in real world outdoor scenarios other variables, like global illumination or location of dynamic objects, might vary on an equal or slower time scale than the position of the robot. To prevent encoding of said variables we propose to restructure the temporal order of training samples based on loop closures in the trajectory. Every time the robot passes by a previously visited place, former recorded images are re-inserted to increase temporal variation of environmental variables. Hence, it is a feedback signal enforcing the model to produce similar outputs due to its slowness objective. Experiments in a simulated outdoor environment demonstrate increased robustness especially for changing lighting conditions.

Keywords: Self-localization, Slow feature analysis, Long-term robustness, Loop closure detection, Omnidirectional vision

1 Introduction

Self-localization is a prerequisite for autonomous mobile robots executing tasks in a spatial environment. The problem of simultaneous localization and mapping (SLAM) has been studied extensively. Recent geometric approaches estimate the motion of the camera and the depth of the scene from consecutive frames based on feature correspondences or semi-direct image alignment [1, 2]. Despite the impressive results demonstrated by these approaches, visual localization and mapping in long-term outdoor scenarios remains a challenging problem due to different lighting conditions, cast shadows, dynamic elements or seasonal effects. The model for SFA-localization is based on the principle of slowness learning [3–5]. It is inspired by the observation that primary sensory signals, like the values of individual pixel, usually change on a faster timescale than the embedded high level information, such as the position of observed objects. Extracting slowly varying features from quickly changing sensory signals thus

should yield the underlying abstract information. It has been shown that a hierarchical SFA-network can model place and head direction cells [6], reflecting spatial information in the brain of rodents, and enable localization of a mobile robot in outdoor environments [7] by directly processing the high dimensional visual input. After an offline learning phase localization is instantaneous and absolute since SFA-outputs are computed from a single sample. Despite the good localization results the learned representation are likely only valid for short periods of time when the image statistics are similar to the training. Here we extend the model using loop closures in the trajectory to restructure the training data for improved robustness. Images from loop closures, representing the same place under different environmental conditions, are re-inserted in the temporally ordered image sequence. This increases temporal variation of environmental effects and is a feedback signal for the SFA-model that has to find functions producing a similar output due to its slowness objective.

2 Model for SFA Localization

2.1 Slow Feature Analysis and Orientation Invariance

SFA as introduced in [8] transforms a multidimensional time series $\boldsymbol{x}(t)$ to slowly varying output signals. The objective is to find instantaneous scalar input-output functions $g_j(\boldsymbol{x})$ such that the output signals $y_j(t) := g_j(\boldsymbol{x}(t))$ minimize $\Delta(y_j) := \langle \dot{y}_j^2 \rangle_t$ under the constraints $\langle y_j \rangle_t = 0$ (zero mean), $\langle y_j^2 \rangle_t = 1$ (unit variance), $\forall i < j : \langle y_i y_j \rangle_t = 0$ (decorrelation and order) with $\langle \cdot \rangle_t$ and $\dot{y}$ indicating temporal averaging and the derivative of y, respectively. The Δ-value is a measure of the temporal slowness of the signal $y_j(t)$, so small Δ-values indicate slowly varying signals. The constraints avoid the trivial constant solution and ensure that different functions g code for different aspects of the input. We use the MDP [9] implementation of SFA, which is based on solving a generalized eigenvalue problem. For the task of self-localization, we want to find functions that encode the robot‘s position as slowly varying features and are invariant w.r.t. its orientation. The information encoded in the learned slow features depends on the statistics of the training data. To achieve orientation invariance, the orientation of the robot has to change on a faster timescale than its position. We use an omnidirectional mirror as a feasible realization to simulate additional robot rotation, by shifting a sliding window over the periodic panoramic images.

2.2 Self-localization and Learned Representations

The high dimensional visual input is processed by a hierarchical network made of several converging layers. In our standard approach [7] layers are trained subsequently with all training images in temporal order of their recording. A single node per layer is trained with stimuli from all node locations in its layer and replicated throughout the layer after training. The model is shown in Fig. 1. The learned spatial representations are analyzed qualitatively by plotting the

color coded eight slowest SFA-outputs $f_{1...8}$ over all positions. In case of strong spatial coding a clear gradient along the coordinate axis should be visible in the spatial firing maps of the first two units. Maps of higher units show a mixture of the first two units or higher modes. For a quantitative analysis we compute a regression function from the SFA-outputs to the ground truth coordinates and apply it to slow features extracted from images of a separate test set.

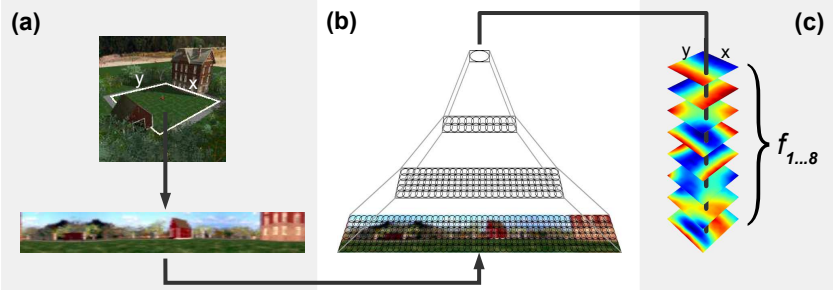


Fig. 1: **Model architecture**. (a) The views from a certain position (x, y) are steadily captured and transformed to a panoramic view. (b) The view is processed by the network. Each layer consists of overlapping SFA-nodes arranged on a regular grid. Each node performs linear SFA for dimensionality reduction followed by a quadratic SFA for slow feature extraction. The output layer is a single node, whose eight slowest outputs $y_j(t)$ are the orientation invariant encoding of the location. (c) The color coded SFA-outputs $f_{1...8}$ over all positions, so-called *spatial firing maps*, ideally show characteristic gradients along the coordinate axes and look the same independent of the specific orientation.

2.3 Extending the Model Using Feedback from Loop Closures

Since the SFA-model processes raw pixel values the learned representations are susceptible to appearance changes of the environment varying on an equal or slower timescale than the position of the robot. We use invariance learning, which is the basis of SFA, to learn representations that are not affected by environmental changes during the training phase. Loop closure detections allow to re-insert images of the same place, with a possibly different appearance, in the temporal sequence of training images thereby increasing the timescale of environmental effects. Further it is a feedback signal for the SFA-model, because its objective is to find functions that produce similar outputs for temporally close training samples.

Loop Closure Detection To validate the feasibility of the approach we first use ground truth information to identify loop closures. A positive match requires that the spatial distance between the match candidates is smaller than a predefined threshold and that there is a minimum temporal gap between them.

In real world scenarios, where no ground truth is available, loop closures can be identified using image information. A common approach is the bag of words (BOW) model describing each image by the occurrences of visual words from a dictionary (e.g. [10]). Here we created a vocabulary of 1500 visual words using Surf-Features [11]. Loop closure matches are then determined comparing the distance between visual word histograms.

Training Using Feedback The SFA-model is trained with the temporally ordered images like in the standard approach. In case of a loop closure match, the past image is aligned to the current one by finding the relative orientation which minimizes the image distance. The aligned image is then re-inserted in the training sequence.

3 Experimental Results

3.1 Experimental Setup

Experiments are conducted in a simulated park like environment covering an area of 16x18. Images from the simulated omnidirectional camera are captured with a resolution of 500x500 pixel and transformed to panoramic views with a size of 600x55 pixel. The training and test trajectory consist of 1773 and 1090 poses that evenly cover the area. Crossings in the training trajectory improve spatial coding of the SFA-model and enable the extended model to get feedback from loop closures. The trajectories and the 62 loop closures determined from ground truth information are illustrated in Fig. 2.

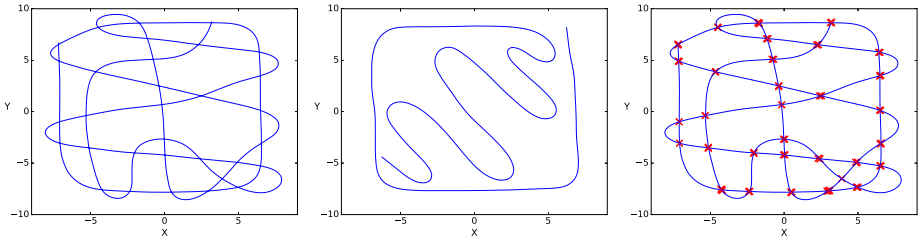


Fig. 2: Left: Training trajectory, Middle: Test trajectory, Right: Loop closures

3.2 Localization in a Static Environment

Initially we compare the standard and the extended model in a static environment to give reference under optimal conditions and to investigate the effect of using feedback from loop closures.

Results Since the feedback only slightly changes the distribution of visited places, resulting representations of both models are nearly identical, leading

to the conclusion that using feedback does not deteriorate performance. Spatial firing maps of the first two units, shown in Fig. 3a and 3c, show clear gradients along the coordinate axis. Units three and four are mixtures of the first two units. Estimated trajectories illustrated in Fig. 3b and 3d are close to the ground truth with mean Euclidean deviations of 0.24 and 0.23, respectively.

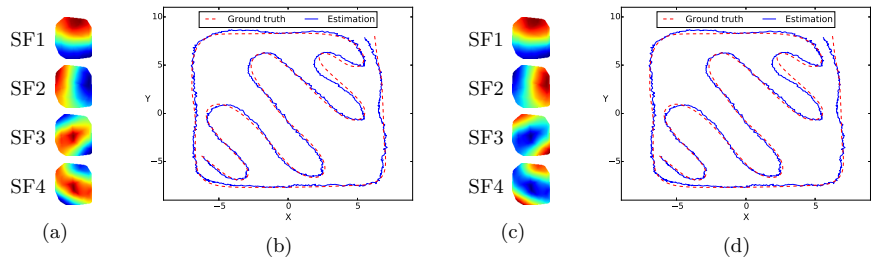


Fig. 3: **Results in the static environment.** Spatial firing maps of the standard (a) and extended model (c) show strong spatial coding. Estimated trajectories of the respective models in (b) and (c) are close to the ground truth.

3.3 Localization with Changing Light

In this experiment we investigate the effect of changing lighting conditions on the localization performance of the standard model and validate the feasibility of the feedback mechanism for improved robustness. The intensity of an artificial point light is increased over the training run leading to non-trivial illumination changes. Light intensity is thus the slowest varying latent variable embedded in the image statistics. Training images illustrating the effect are shown in Fig. 4.

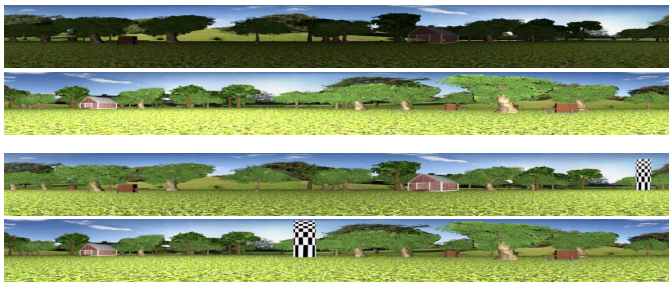


Fig. 4: **Environmental effects.** First and last image of the training sequences. Top: The effect on the appearance of increasing light intensity over the run, Bottom: A textured cylinder is moved along a circle around the training area.

Results The quality of the spatial representations learned by the standard model is clearly deteriorated by the changing light intensity. Spatial coding is not observable in the spatial firing maps shown in Fig. 5a, while at least some position information is contained in the SFA-units since the estimated trajectory is not random (see Fig. 5b). The mean Euclidean deviation from the ground truth is 2.4. Using feedback from the loop closures enables the SFA-model to become more invariant against changing light intensity. Spatial firing maps illustrated in Fig. 5c show a clear gradient along the coordinate axis for SFA-units one and two, while units three and four are mixtures of the first two units. The mean Euclidean deviation from ground truth amounts to 0.49. The estimated trajectory can be seen in Fig. 5d.

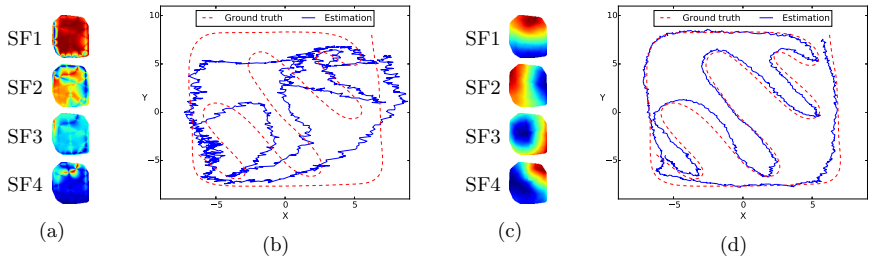
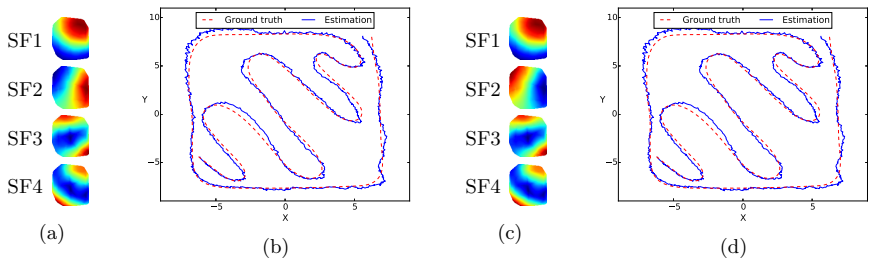


Fig. 5: **Results with changing light.** (a) Spatial firing maps from the standard SFA-model show no spatial coding. (b) Localization performance is deteriorated but not random which indicates at least weak position coding. (c) Characteristic gradients along the coordinate axis in the spatial firing maps of first two units from the extended model suggest good spatial coding. (d) Localization accuracy clearly improves with the extended model.

3.4 Localization With a Dynamic Object

In this experiment we investigate the effect of a dynamic object. A textured cylinder is moved along a circle around the training area performing one circumnavigation during the training phase so that its location is the slowest changing variable. Fig. 4 shows the first and last image containing the dynamic object. **Results** The effect of the dynamic object on the resulting representations is not as big as expected. Spatial firing maps of the first two SFA-units from both models, shown in Fig. 6a and 6c, show gradients along the coordinate axis. Accuracy of the estimated trajectories is only slightly worse than in the static environment as both models achieve a mean Euclidean deviation of 0.29. Estimated trajectories of both models are shown in Fig. 6b and 6d. The dynamic object seems to produce local noise only but no high level information about its position is encoded in the SFA-units.



(a) (b) (c) (d)

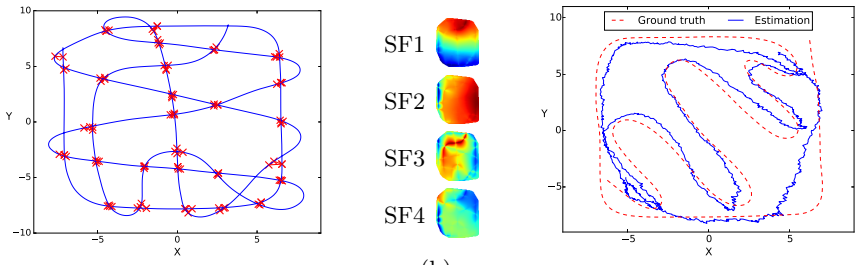
Fig. 6: **Results with a dynamic object.** Spatial firing maps from the standard model (a) and maps of the model using feedback (c) are nearly identical showing clear gradients along the coordinate axis. Estimated trajectory of the standard model (b) and the extended model (d) are close to the ground truth.

3.5 Localization Using Feedback from BOW Loop Closures

Ground truth loop closures used in the previous experiments have a mean Euclidean distance of 0.06 between match candidates. In this experiment we use a bag of words model for loop closure detection. Defining 0.1 as the maximum Euclidean distance for a positive match results in a mean average precision of 0.52. The 54 accepted matches with a mean Euclidean distance of 0.27 are depicted in Fig. 7a. The experiment is performed on the data set featuring changing light intensity since the effect of using the feedback was clearly visible.
Results The imprecise loop closures interfere the quality of the resulting representations. The effect of the changing light intensity is visible in the spatial firing maps in Fig. 7b. First two units show a gradient along the coordinate axis but are not as smooth as in the experiment using ground truth matches. Units three and four are clearly influenced by the changing environmental variable. The localization accuracy is interfered accordingly with a mean Euclidean deviation of 0.9. Nonetheless, this is a big improvement compared to the standard model with a mean deviation of 2.4. The estimated trajectory is shown in Fig. 7c.

4 Conclusion

We presented an extension of the biologically motivated model for SFA-localization using feedback from loop closures to improve robustness. Re-inserting images of the same place from the past in the temporally ordered image stream increases variation of environmental effects and is a feedback signal for the SFA-learning algorithm regarding its slowness objective. We have shown that feedback from loop closures improves robustness especially for changing lighting conditions. Experiments with loop closure matches from a BOW-model suggest the applicability of the model in real world scenarios. An elaborate solution to deal with imprecise loop closures could be the use of a weighted SFA-formulation, as described in [12, 13], with training samples organized in a graph where the connecting edges represent their similarity regarding the labels.



(a) (b) (c)

Fig. 7: **Results with changing light using feedback from BOW loop closures.** (a) Loop closures determined by matching visual word histograms. (b) Spatial firing maps suggest position coding in the first two units while units three and four show the influence of changing light intensity. (c) Localization performance clearly surpasses the standard model while deviations are larger compared to the model using ground truth loop closures.

References

1. Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: ORB-SLAM: a versatile and accurate monocular SLAM system. CoRR abs/1502.00956 (2015)
2. Engel, J., Schöps, T., Cremers, D.: LSD-SLAM: large-scale direct monocular SLAM. In: ECCV, Switzerland. pp. 834–849 (2014)
3. Földiák, P.: Learning invariance from transformation sequences. Neural Computation 3(2), 194–200 (1991)
4. Stone, J., Bray, A.: A learning rule for extracting spatio-temporal invariances. Network: Computation in Neural Systems 6(3), 429–436 (1995)
5. Körding, K.P., Kayser, C., Einhäuser, W., König, P.: How are complex cell properties adapted to the statistics of natural stimuli? J. Neurophysiol. 91(1), 206–212 (2004)
6. Franzius, M., Sprekeler, H., Wiskott, L.: Slowness and sparseness lead to place, head-direction, and spatial-view cells. PLoS Computational Biology 3(8) (2007)
7. Metka, B., Franzius, M., Bauer-Wersing, U.: Outdoor self-localization of a mobile robot using slow feature analysis. In: ICONIP, Korea. pp. 249–256 (2013)
8. Wiskott, L., Sejnowski, T.: Slow feature analysis: Unsupervised learning of invariances. Neural Computation 14(4), 715–770 (2002)
9. Zito, T., Wilbert, N., Wiskott, L., Berkes, P.: Modular toolkit for data processing (mdp): a python data processing framework. Front. Neuroinform. 2(8) (2009)
10. Cummins, M., Newman, P.: Highly scalable appearance-only SLAM - FAB-MAP 2.0. In: RSS, USA (2009)
11. Bay, H., Tuytelaars, T., Gool, L.J.V.: SURF: speeded up robust features. In: ECCV, Austria. pp. 404–417 (2006)
12. Escalante, A.N., Wiskott, L.: How to solve classification and regression problems on high-dimensional data with a supervised extension of slow feature analysis. Journal of Machine Learning Research 14(1), 3683–3719 (2013)
13. Escalante, A.N., Wiskott, L.: Improved graph-based SFA: information preservation complements the slowness principle. CoRR abs/1601.03945 (2016)