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

Robust Outdoor Self-localization In Changing Environments

Muhammad Haris, Mathias Franzius, Ute Bauer-Wersing

2019

Preprint:

This is an accepted article published in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019). The final authenticated version is available online at: https://doi.org/[DOI not available]

Robust Outdoor Self-localization In Changing Environments

Muhammad Haris[1] and Mathias Franzius[2] and Ute Bauer-Wersing[1]

***Abstract*— In outdoor scenarios changing conditions (e.g., seasonal, weather and lighting effects) have a substantial impact on the appearance of a scene, which often prevents successful visual localization. The application of an unsupervised Slow Feature Analysis (SFA) on the images captured by an autonomous robot enables self-localization from a single image. However, changes occurring during the training phase or over a more extended period can affect the learned representations. To address the problem, we propose to join long-term recordings from an outdoor environment based on their position correspondences. The established hierarchical model trained on raw images performs well, but as an extension, we extract Fourier components of the views and use that for learning of spatial representations, which reduces the computation time and makes it adequate to run on an ARM embedded system. We present the experimental results from a simulated environment and real-world outdoor recordings collected over a full year, which has effects like different day time, weather, seasons and dynamic objects. Results show an increasing invariance w.r.t. changing conditions over time, thus an outdoor robot can improve its localization performance during operation.**

I. INTRODUCTION

Nowadays service robots are becoming an attractive solution to perform various daily life tasks, which provides a way to decrease human efforts. The ability of an autonomous mobile agent to locate itself in unknown environments is an essential criterion to implement intelligent behavior, and for that, it needs an internal representation of its surrounding. Laser-based SLAM in 2D, flat environments solve the problem of simultaneous localization and mapping (SLAM) [1], but the high cost limits its use in domestic applications. On the other hand, vision-based SLAM (vSLAM) provides an alternative to address the problem as cameras are increasingly inexpensive and small. Although recent advances in the field have shown impressive results in the mapping of large-scale environments [2], [3], the long-term outdoor operation in unconstrained environments is still an active research area [4], [5]. Vision-based autonomous mobile robots operating outdoors over a long time have to deal with changing environments. These changes may come from different lighting conditions, weather, seasonal shift or due to a structural change. This problem poses a great challenge for robots that aim to perform long-term self-localization. This work focuses on the efficient learning of spatial representations that become invariant to such changes, which would help to achieve long-term stability w.r.t. environmental conditions.

[1]Muhammad Haris and Ute Bauer-Wersing are with the Faculty of Computer Science and Engineering, Frankfurt University of Applied Sciences, 60318 Frankfurt, Germany muhammad.haris@fb2.fra-uas.de

[2]Mathias Franzius is with the Honda Research Institute Europe GmbH, 63073 Offenbach, Germany

Various animals, e.g., rodents show exceptional navigation capabilities in natural environments. Their brain does not have direct access to spatial information; instead, it depends on sensory signals from eyes to extract the behaviorally relevant information, i.e., position or orientation of an animal in space. The hippocampal region of the rodent's brain has different cell types, which encode spatial information. For instance, Place Cells fire if the animal is present in a specific part of the environment, independent of its head-direction [6]. On the other hand, Head Direction Cells are active when the animal is looking in a particular direction while being invariant to its position [7]. In [8], the authors show that firing activity of Place or Head Direction Cells strongly depends on visual input. Earlier work [9] demonstrates that a hierarchical model trained in an unsupervised way with raw visual input as perceived by a virtual rat reproduces Place or Head Direction Cells characteristics. The model uses the concept of temporal slowness to learn relevant information. The core idea behind slowness learning is that primary sensory signals (here: pixel values in a video) typically vary on a faster time scale compared to the significant information, e.g., an animal's position in space. This observation has led to the concept of slowness learning [10], [11], [12]. Our work utilizes the Slow Feature Analysis (SFA) implementation [13] for learning spatial representations of an environment.
A theoretical study [9] of the hierarchical SFA-model used for the task of self-localization in open space shows that in slowness learning, the learned spatial representation primarily depends on the movement statistics of the animal during the training phase. Thus, if the position varies relatively slower than the head-direction during the mapping of an environment, the slowest features learned by SFA will code for the position. Previous work [14] has successfully demonstrated SFA based self-localization in real-world outdoor environments and established a system that learns instantaneous orientation invariant representations of the robot's position in an unsupervised learning process. Moreover, the work achieved equal or better results on a variety of test scenarios compared to state of the art visual SLAM methods [3], [15]. Since the learned representations of location are strongly affected by environmental conditions their practical use, especially in outdoor scenarios is limited. To deal with the issue, the authors proposed a method [16] to re-insert images in the training sequence based on loop closure events to learn an invariance w.r.t. environmental changes during training. Images from loop closures represent the same place in different conditions. The re-insertion of these views changes the robot's perceived temporal input statistics, which allows learning an invariance to such changes [17]. The pro-

posed method enables mitigating the effects of environmental changes occurring during a single training run. Despite the improved robustness, the learned representations will only be valid if image statistics of the training and testing phase are quite similar. For this reason, we extend this approach to long-term recordings from the same trajectory to achieve long-term stability w.r.t. environmental changes.

In this work, we first use high-dimensional panoramic views from long-term recordings to train the four-layer hierarchical SFA network with the same parameter settings as in [14]. By applying the proposed training scheme (see Section III), the slowest features learned by the SFA model should ideally code for the robot's position while being invariant to condition changes. Although the learning of spatial representations based on raw images performs well it has two limitations. Firstly, for many realistic robot trajectories, learned representations may contain mixtures of position and orientation, requiring additional steps to overcome [9], [14]. Secondly, the training and execution times of hierarchical SFA are too slow for a low-power embedded system (see Section V). Therefore, as an alternative to using raw images, we extract Fourier components of panoramic views and use that representation to learn spatial encoding. This preprocessing step removes the orientation dependency (see Section III). Further, it allows using a two-step SFA instead of the hierarchical network, which notably decreases the computation time of training and execution phases. Hence, it makes the system worthwhile for implementing it on a real robot equipped with low-cost hardware.

Section II reviews the related work focusing on performing reliable localization in changing conditions. Section III gives a mathematical definition of Slow Feature Analysis. Further, it presents the reordering of the training sequence for learning condition invariance and Fourier feature extraction. Section IV shows the experimental results for self-localization in changing conditions for simulated and real-world data. Section V concludes the work and outlines the follow-up work.

II. RELATED WORK

There are a variety of methods, which aim to achieve long-term stability w.r.t. changing conditions for robust self-localization. Several authors have addressed the problem partially by focusing only on illumination changes. The approaches include, for example, an active exposure control method for visual odometry in high dynamic range environments [4], image transformation into lighting-invariant color space [5] and visual feature point descriptors [18]. In [19], the system does not build a single monolithic map that represents all the observations of a workspace; instead, it creates a composite representation from multiple runs in a workspace to capture the diversity of varying conditions. Despite the system's performance, memory demand and map complexity increase over time. The feature-finding algorithms [20] are well known for performing the task of place recognition but may fail to deal with extreme visual change. The use of image sequence matching has shown significant improvement in visual localization even with drastic appearance changes [21]. Initially, the images are projected to more robust representation by down-sampling and patch normalization steps. Finally, the approach calculates the sequence with minimum cost instead of estimating a single global match. Although the results are impressive, this approach assumes the same route for each run, which makes it less attractive for performing localization in open field scenarios. Other approaches [22], [23] use features from pre-trained deep convolutional neural networks for the task of place recognition. Features extracted from different layers are invariant w.r.t. to viewpoint and condition changes. However, the computation and matching of the high dimensional features are computationally expensive, which may not be suitable for real-time operation on a mobile robot. In [24], the authors have used PoseNet [25], which is trained end-to-end to estimate the camera's six DOF pose from a single monocular image. However, it is also computationally expensive and requires ground truth positional data.

III. METHODS

A. Slow Feature Analysis

Slow Feature Analysis (SFA) is an unsupervised learning algorithm, which extracts slowly varying features from a quickly varying input signal. SFA has been successfully applied to self-organization of complex-cell receptive fields, invariant object recognition, the self-organization of the place-cells and to nonlinear blind source separation [26]. From a mathematical point of view, Slow Feature Analysis transforms a set of time varying input signals $\mathbf{x}(t)$ into a set of slowly varying output signals $\mathbf{s}(t)$. The optimization objective is to find functions $g_j(\mathbf{x})$ such that the output signals [13]

$$s_j(t) := g_j(\mathbf{x}(t))$$

minimize

$$\Delta(s_j) := \langle \dot{s}_j^2 \rangle_t$$

under the constraints

$$\begin{aligned} \langle s_j \rangle_t &= 0 \text{ (zero mean)}, \\ \langle s_j^2 \rangle_t &= 1 \text{ (unit variance)}, \\ \forall i < j : \langle s_i s_j \rangle_t &= 0 \text{ (decorrelation and order)} \end{aligned}$$

The $\langle \cdot \rangle_t$ and $\dot{s}$ represents time averaging and derivative of s, respectively. The Δ-value defines the temporal variation of the output signal $s_j(t)$ and its minimization is the objective function. A lower value indicates less variation of the signal over time thus means slowly varying signals. The constraints avoid the trivial constant solution, $s_j(t) = constant$ and ensure that different functions g encodes different aspects of the input. We use the MDP [27] implementation of SFA, which is based on solving a generalized eigenvalue problem.

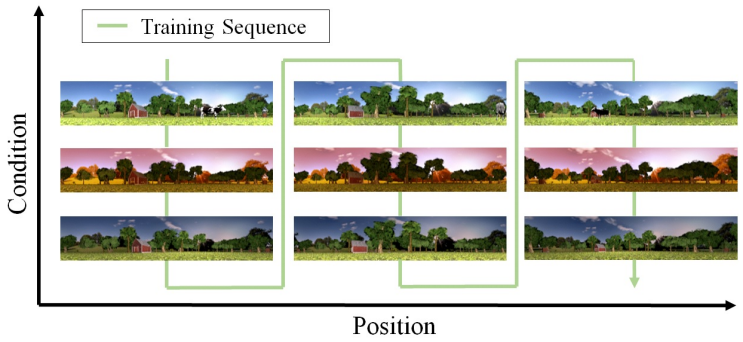


Fig. 1. **Restructuring scheme to learn invariance w.r.t. varying conditions.** The training sequence includes images from the same trajectory in different environmental conditions. The joining of long-term recordings based on position allows creating a training sequence where environmental conditions change faster than the robot's position, which will enable SFA to learn invariance to such changes.

B. Learning Environmental Invariance with SFA

Unsupervised learning of spatial representation with Slow Feature Analysis (SFA) allows self-localization by processing the views captured during a training phase. The variable of interest for the task of self-localization is the robot's position. If during training, the robot's position changes slowly relative to other variables (for instance the robot's orientation and environmental changes), the robot's position will be the slowest feature learned by SFA.

In a controlled setting with no environmental changes, the resulting SFA-outputs code for the position or orientation of the robot depending on the movement statistics during training [9]. However, in real-world outdoor scenarios, it is quite rare to have unvarying environmental conditions. If these changes occur on a slower or same timescale than the robot's position, the resulting representation may encode them as slowly varying features, which can prevent successful localization. Previous work [16] uses a data-driven approach to learn invariance w.r.t. short-term environmental changes during a single training run. The proposal was to restructure the temporal order of the training sequence based on loop closures in a trajectory. Thus, every time the robot revisits a place, previously recorded images are re-inserted in the training sequence. Views of the same place in different conditions thus appear temporally close. Hence, it serves as a feedback signal, which enforces the SFA model to produce similar outputs at similar locations due to its slowness objective. This approach has shown good localization results but will fail if the image statistics of the training and test set are entirely different. Here, we extend the approach to long-term recordings from the same trajectory. For each recording, a robot automatically traverses a fixed closed loop trajectory. During traversal, it stores views of the environment and odometry information. We reordered the recordings by establishing position correspondences between them using the odometry. Based on this association, it is possible to combine the images of the same place in different conditions followed by inserting them into a training sequence and then proceeding to the next position in a trajectory. It will cause the environmental condition to vary faster than the position of the robot in the training data, which will enable the learning of invariant representations w.r.t. environmental conditions while keeping spatial position encoding. Fig. 1 illustrates the organization of the training data.

C. Fourier Feature Extraction

Representations learned with SFA strongly depend on temporal statistics and much less on modality or preprocessing [9]. Extracting Fourier components acts here as a compression that preserves temporal statistics and distinct sensory representations for each position. As a first step for extracting Fourier components, we project omnidirectional views captured during the training phase to panoramic images. The second step is to perform row-wise Fourier series expansion and store only the magnitude part corresponding to the lowest 15 frequency components. Please note that a different perspective of the same location usually degrades localization performance. However, we can obtain orientation invariance for each location by storing the magnitude part of Fourier components since it is not dependent on robot direction, as shown in [28], [29]. The compact Fourier representation obtained for each image by the preprocessing step is used to learn SFA representations. The inclusion of Fourier preprocessing omits the need for hierarchical SFA-network as in this case, the learning phase only consists of two steps; the first one reduces the dimensionality by using a linear SFA while the second extract non-linear slow features using a quadratic SFA. The input and output dimensionality for the first step are 450 and 20, respectively. The input and output dimensionality for the second step are 20 and 8, respectively. The eight SFA-output units $s_{1...8}$ represent the slowest features learned by SFA. After the training phase, the computation of slow features is instantaneous, i.e., only need a single image to compute the model output.

IV. EXPERIMENTS

We apply the reordering scheme as mentioned earlier to the synthetic recordings generated from a virtual reality simulator. Later we apply it to real-world recordings gathered from an outdoor environment.

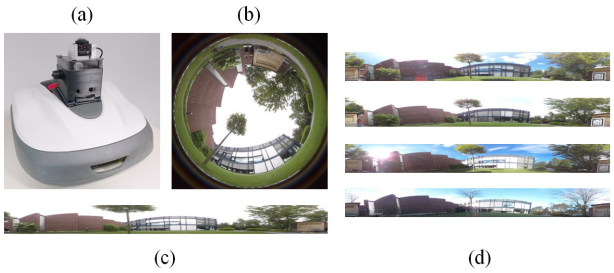


Fig. 2. (a) A mower robot with an omnidirectional camera used for the experiments. (b) An omnidirectional image. (c) The corresponding panoramic image. (d) Same place under different conditions.

A. Simulated Environment

To simulate an outdoor environment, we used Blender[1] software to generate images. A virtual robot traverses a trajectory of area 15×15 meter and captures omnidirectional views of a scene. The omnidirectional images have the size of 300×300 pixels. As a preprocessing step, we project the images to panoramic views having 600×60 dimension. The data set consists of 10 recordings generated by random variation of lighting parameters. The lighting parameters include energy $\subset [3,8]$, the y-coordinate of the light source $\subset [-10,10]$m and the intensity of the red channel $\subset [0.5,1]$. Each image set has 279 panoramic images with a size of 600×60 pixels. As discussed earlier, we reordered the long-term recordings based on position correspondences such that the environmental condition varied faster than the robot's position in the training sequence (Fig. 1). The model uses long-term data for the training phase in an incremental way. The model trained with n different conditions uses the next unseen condition $n+1$ in the data set for performance evaluation. We use the trained model and process the training images to get an SFA-matrix, where each row corresponds to the first eight slowest features $s_{1\ldots8}$ for each location (x,y). Afterwards, we use this matrix to train a separate regression function for x and y metric ground truth positions. Similarly, we obtain the SFA-matrix for the test data and use the learned regression function to predict the test set locations (x',y'). Finally, we compute the mean absolute error between the predicted (x',y') and actual test positions (x,y). Please note that SFA is unsupervised, ground truth data used here serves the purpose of metric evaluation.

For validation purpose, we repeated the same procedure with 15 random permutations of the image sets. Fig. 3 shows the mean localization performance of the experiments and an estimated trajectory for the 10th set using 9 data sets for training. Mean absolute error for the last test set is 0.5 and 0.4 meters for hierarchical SFA on images and SFA on Fourier preprocessed images, respectively. Using the representations learned in a single condition has a prohibitively large error in a different condition but it decreases quickly for more data sets. Thus, the results have shown that it is possible to learn an increasingly invariant representation of the environment.

[1]https://www.blender.org/

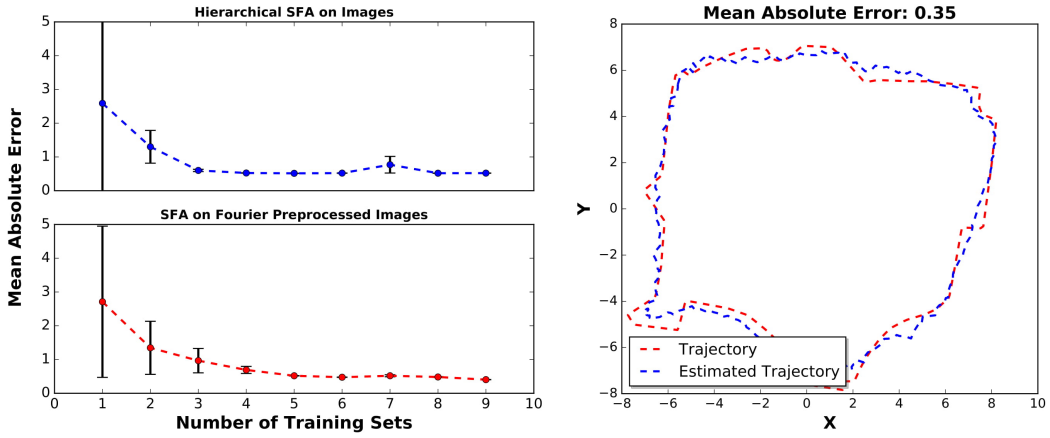
B. Real-world outdoor environment

This section demonstrates the application of the proposed strategy on real-world outdoor data. We use a modified lawn mower robot that has a fish-eye camera mounted on top of it to capture views of an environment. It captures omnidirectional images of size 2880×2880 pixels. Fig. 2 shows the robot, an omnidirectional view captured by the camera, its associated panoramic image, and some images from our dataset. The data set has 15 long-term recordings collected over a full year, e.g., 12 months. These recordings have environmental effects like different daytimes, weather, seasons and dynamic obstacles. During the data collection phase for each recording, the mower uses the standard wire guidance technology, where it follows a wire loop buried in the ground. As the start and end positions, as well as the orientation within the base station, are known it is possible to use the odometry estimation from the robot to obtain precise position estimates for metric train and test labels [30]. The garden area enclosed by border wire is 15×9 meter.

The proposed strategy requires to combine the long-term sets based on their nearest neighbor position correspondences. After establishing the association, each data set consists of 799 panoramic images with a resolution of 600×60 pixels. Here, we performed the same series of experiments as in the simulator environment. Fig. 4 shows the mean localization performance of the experiments and an estimated trajectory for the 15th set using 14 data sets. Mean absolute error for the last test set is 0.7 and 1.1 meters for hierarchical SFA on images and SFA on Fourier preprocessed images, respectively. As to be expected, the use of a single training set is not sufficient to perform localization in a different environment as the mean error is quite high. However, adding successive data sets to the training sequence quickly and noticeably reduces the error.

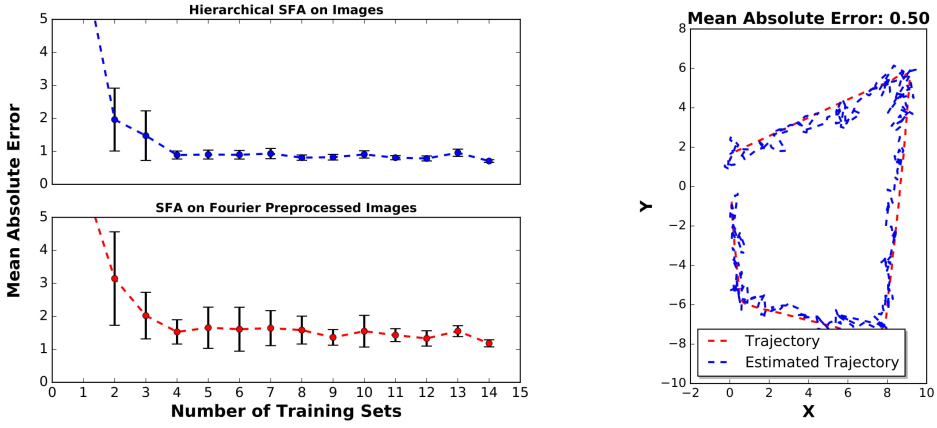
V. DISCUSSION

The goal of this work was to address the problem of localization in changing conditions. Learning an invariance to such changes allows a robot to localize itself robustly over a long period from single images. Although the appearance of an environment increasingly changes with seasons, localization accuracy improves over time. We have joined long-term data sets based on their position correspondences such that environment conditions vary faster than the robot's position in the training sequence. It allows SFA to encode the robot's position while learning invariance to these changes due to its slowness objective. We have applied the reordering scheme initially to a simulated outdoor environment and then to the data collected from a real-world outdoor environment throughout a year. As an alternative to learning on raw images, we applied the Fourier preprocessing step, which has removed the dependency of learning orientation invariance and hierarchical SFA processing. This approach drastically reduces the computation time. On a standard PC, a full training session takes 100 s on raw images, and localization from a single image takes 165 ms. With the Fourier-based approach, the respective times are 2.85 s and 1.1 ms, which



(a) (b)

Fig. 3. **Simulated environment:** (a) Mean and variance localization performance for 15 random permutations of 10 simulated recordings for learning with raw images (Top) and Fourier-preprocessed images (Bottom). Each point indicates localization error for the next unseen condition. Both plots show similar characteristics, an increasingly invariant representation of the environment is possible to learn by adding training sets, which have a high variation w.r.t. environmental conditions. (b) Estimated trajectory for an unseen set (blue) using 9 image sets with a mean absolute error of 0.35 meter using the Fourier-based model.



(a) (b)

Fig. 4. **Real-World environment:** (a) Mean and variance localization performance for 15 random permutations of 15 real-world recordings for learning with raw images (Top) and Fourier-preprocessed images (Bottom). Each point indicates localization error for the next unseen condition. The resulting plots show a decrease in localization error with adding the number of training sets with different conditions. (b) Estimated trajectory for an unseen set (blue) using 14 image sets with a mean absolute error of 0.50 meter using the hierarchical SFA model.

is a significant improvement. Similarly, on a i.MX6 ARM board, single core 800 MHz, the localization from a raw image takes 2.85 s, and Fourier preprocessed image takes 20 ms. However, this improvement comes with a small loss in localization accuracy as the mean error for image-based learning is 0.7 meter and for Fourier-based learning is 1.1 meters for the real-world outdoor environment. To conclude, we can cope with long-term instability from a variety of sources (e.g., lighting, weather, seasons) with a straightforward model, which is independent of any feature descriptors. Further, we can do it very efficiently by using Fourier representations. The experiments have shown that an agent can autonomously learn an invariant representation of the environment. The current SFA-model uses visual input to localize; however, preliminary data shows an additional 30% improvement in accuracy by adding a Kalman filter for odometry integration. Preliminary data also suggest that the representations learned only on the area wire directly generalize to localization in the full field.

References

[1] D. Hahnel, W. Burgard, D. Fox, and S. Thrun, "An efficient fastslam algorithm for generating maps of large-scale cyclic environments from raw laser range measurements," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, vol. 1, Oct 2003, pp. 206–211 vol.1.

[2] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *Int. J. Rob. Res.*, vol. 27, no. 6, pp. 647–665, June 2008. [Online]. Available: http://dx.doi.org/10.1177/0278364908090961

[3] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *CoRR*, vol. abs/1502.00956, 2015. [Online]. Available: http://arxiv.org/abs/1502.00956

[4] Z. Zhang, C. Forster, and D. Scaramuzza, "Active exposure control for robust visual odometry in hdr environments," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 3894–3901.

[5] C. McManus, W. Churchill, W. Maddern, A. D. Stewart, and P. Newman, "Shady dealings: Robust, long-term visual localisation using illumination invariance," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 901–906.

[6] J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat," *Brain Research*, vol. 34, no. 1, pp. 171 – 175, 1971.

[7] J. Taube, R. Muller, and J. Ranck, Jr, "Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 10, pp. 420–35, 03 1990.

[8] K. J. Jeffery and J. M. O'Keefe, "Learned interaction of visual and idiothetic cues in the control of place field orientation," *Experimental Brain Research*, vol. 127, pp. 151–161, 1999.

[9] M. Franzius, H. Sprekeler, and L. Wiskott, "Slowness and sparseness lead to place, head-direction, and spatial-view cells," *PLoS Computational Biology*, vol. 3, no. 8, 2007.

[10] P. Földiák, "Learning invariance from transformation sequences," *Neural Comput.*, vol. 3, no. 2, pp. 194–200, June 1991. [Online]. Available: http://dx.doi.org/10.1162/neco.1991.3.2.194

[11] J. STONE, "A learning rule for extracting spatiotemporal invariances," *Network: Computation in Neural Systems*, vol. 6, pp. 429–436, 1995. [Online]. Available: https://ci.nii.ac.jp/naid/80008541560/en/

[12] L. Wiskott, "Learning invariance manifolds," in *ICANN 98*, L. Niklasson, M. Bodén, and T. Ziemke, Eds. London: Springer London, 1998, pp. 555–560.

[13] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: Unsupervised learning of invariances," *Neural Comput.*, vol. 14, no. 4, pp. 715–770, Apr. 2002. [Online]. Available: http://dx.doi.org/10.1162/089976602317318938

[14] B. Metka, M. Franzius, and U. Bauer-Wersing, "Bio-inspired visual self-localization in real world scenarios using slow feature analysis," *PLOS ONE*, vol. 13, no. 9, pp. 1–18, 09 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0203994

[15] J. Engel, T. Schöps, and D. Cremers, "Lsd-slam: Large-scale direct monocular slam," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 834–849.

[16] B. Metka, M. Franzius, and U. Bauer-Wersing, "Improving robustness of slow feature analysis based localization using loop closure events," in *Artificial Neural Networks and Machine Learning – ICANN 2016*, A. E. Villa, P. Masulli, and A. J. Pons Rivero, Eds. Cham: Springer International Publishing, 2016, pp. 489–496.

[17] A. N. Escalante-B. and L. Wiskott, "Improved graph-based SFA: information preservation complements the slowness principle," *CoRR*, vol. abs/1601.03945, 2016. [Online]. Available: http://arxiv.org/abs/1601.03945

[18] N. Carlevaris-Bianco and R. M. Eustice, "Learning visual feature descriptors for dynamic lighting conditions," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2014, pp. 2769–2776.

[19] W. Churchill and P. Newman, "Experience-based navigation for long-term localisation," *The International Journal of Robotics Research*, vol. 32, no. 14, pp. 1645–1661, 2013. [Online]. Available: https://doi.org/10.1177/0278364913499193

[20] C. Valgren and A. J. Lilienthal, "Sift, surf & seasons: Appearance-based long-term localization in outdoor environments," *Robot. Auton. Syst.*, vol. 58, no. 2, pp. 149–156, Feb. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.robot.2009.09.010

[21] M. J. Milford and G. F. Wyeth, "Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights," in *2012 IEEE International Conference on Robotics and Automation*, May 2012, pp. 1643–1649.

[22] N. Sünderhauf, F. Dayoub, S. Shirazi, B. Upcroft, and M. Milford, "On the performance of convnet features for place recognition," *CoRR*, vol. abs/1501.04158, 2015. [Online]. Available: http://arxiv.org/abs/1501.04158

[23] N. Sünderhauf, S. Shirazi, A. Jacobson, F. Dayoub, E. Pepperell, B. Upcroft, and M. Milford, "Place recognition with convnet landmarks: Viewpoint-robust, condition-robust, training-free," in *Robotics: Science and Systems*, Auditorium Antonianum, Rome, July 2015. [Online]. Available: https://eprints.qut.edu.au/84931/

[24] A. Kendall and R. Cipolla, "Geometric loss functions for camera pose regression with deep learning," *CoRR*, vol. abs/1704.00390, 2017. [Online]. Available: http://arxiv.org/abs/1704.00390

[25] A. Kendall, M. Grimes, and R. Cipolla, "Convolutional networks for real-time 6-dof camera relocalization," *CoRR*, vol. abs/1505.07427, 2015. [Online]. Available: http://arxiv.org/abs/1505.07427

[26] L. Wiskott, P. Berkes, M. Franzius, H. Sprekeler, and N. Wilbert, "Slow feature analysis," *Scholarpedia*, vol. 6, no. 4, p. 5282, 2011, revision #137965.

[27] T. Zito, N. Wilbert, L. Wiskott, and P. Berkes, "Modular toolkit for data processing (mdp): a python data processing framework," *Frontiers in Neuroinformatics*, vol. 2, p. 8, 2009.

[28] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro, "Image-based monte-carlo localisation without a map," in *AI*IA 2003: Advances in Artificial Intelligence*, A. Cappelli and F. Turini, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 423–435.

[29] H. Ishiguro and S. Tsuji, "Image-based memory of environment," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS '96*, vol. 2, Nov 1996, pp. 634–639 vol.2.

[30] N. Einecke, K. Muro, J. Deigmöller, and M. Franzius, "Working area mapping with an autonomous lawn mower," in *Conference on Field and Service Robotics*. Springer, September 2017.