25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10|15 January 2021

Our work has been supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) 17K00361, and (C) 20K12008.

Deep Next-Best-View Planner for Cross-Season Visual Route Classification

Kurauchi Kanya, Tanaka Kanji, University of Fukui, Japan

Abstract -This paper addresses the problem of active visual place recognition (VPR) from a novel perspective of long-term autonomy. In our approach, a next-best-view (NBV) planner plans an optimal action-observation-sequence to maximize the expected cost-performance for a visual route classification task. A difficulty arises from the fact that the NBV planner is trained and tested in different domains (times of day, weather conditions, and seasons). Existing NBV methods may be confused and deteriorated by the domain-shifts, and require significant efforts for adapting them to a new domain. We address this issue by a novel deep convolutional neural network (DNN) -based NBV planner that does not require the adaptation step.

Research Goal

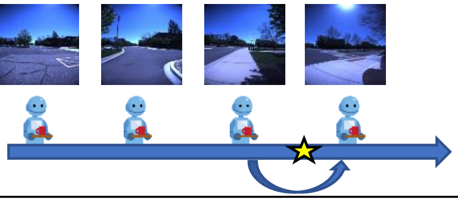
Background : VPR (Long-term visual place recognition)

Goal : classify ego-centric view images into pre-defined place classes

Standard solution : passive setting
the robot's action is determined by a predefined control rule, such as a constant speed motion rule.

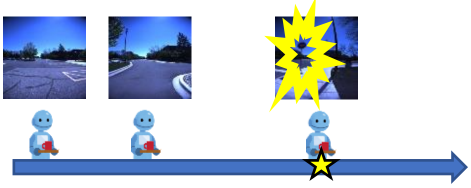
Limitations :

i. viewpoints are not necessarily optimized for the VPR task.
ii. produce an unnecessarily large number of redundant observations.



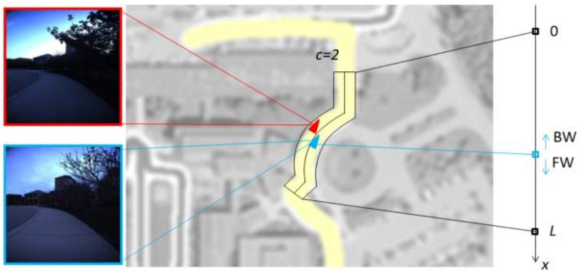
Our approach :

i. active VPR in visual route classification
ii. domain-invariant NBV planner



Proposal: NBV as POMDP (partially observable decision process)

State: s=(x,c)
c: hidden place class
x: viewpoint wrt the place c



25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10|15 January 2021

Our work has been supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) 17K00361, and (C) 20K12008.

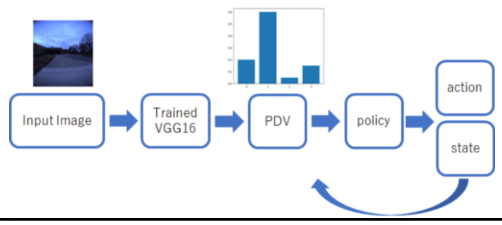
Deep Next-Best-View Planner for Cross-Season Visual Route Classification

Kurauchi Kanya, Tanaka Kanji, University of Fukui, Japan

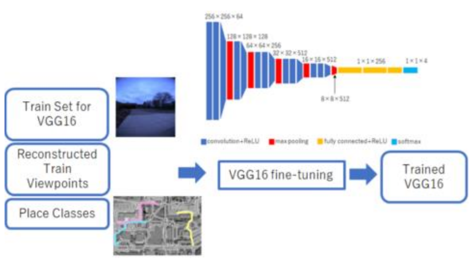
Proposed approach

The goal of an NBV planner is to plan an optimal action-observation-sequence that is expected to maximally improve the cost performance of VPR.

Training VPR

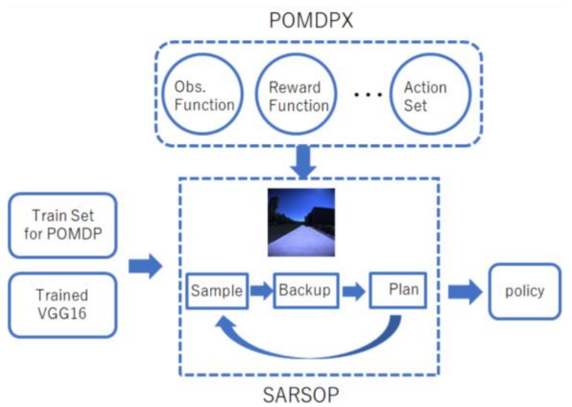


Training planner



Planning NBV

Following the definition by Kaelbling, we define POMDP as a six-tuple (S,A,T,R,W,O)



Result

a. the conventional passive single-view VPR
b. a multi-view VPR
c. the proposed multi-view VPR method

(a) success rate [%]

	ws1	ws2	ws3
single-view	86.7	64.5	77.1
multi-view	84.5	64.5	77.2
proposed	87.8	66.7	82.8

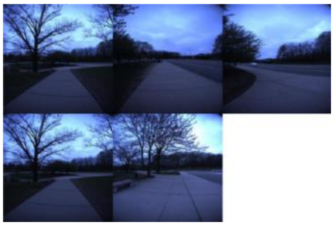
(b) average number of observations

	ws1	ws2	ws3
single-view	1.00	1.00	1.00
multi-view	2.57	2.67	2.66
proposed	2.14	1.99	2.51

(c) average travel distance

	ws1	ws2	ws3
single-view	0.0	0.0	0.0
multi-view	79.2	89.3	78.1
proposed	34.3	34.4	40.0

The figure below shows examples of the image sequences acquired at the planned viewpoints.


Example 1


Example 2

References

[1] S. M. Lowry, N. Sunderhauf, P. Newman, J. J. Leonard, D. D. Cox, P. I. Corke, and M. J. Milford, “Visual place recognition: A survey,” IEEE Trans. Robotics, vol. 32, no. 1, pp. 1–19, 2016.
[2] S. Lowry, N. Sunderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, “Visual place recognition: A survey,” vol. 32, no. 1. IEEE, 2015, pp. 1–19.
[3] A. Kendall, M. Grimes, and R. Cipolla, “Posenet: A convolutional network for real-time 6-dof camera relocalization,” in Proceedings of the IEEE international conference on computer vision, 2015, pp. 2938– 2946.
[4] A. Kendall and R. Cipolla, “Geometric loss functions for camera pose regression with deep learning,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5974–5983.
[5] E. Muller-Budack, K. Pustu-Iren, and R. Ewerth, “Geolocation estimation of photos using a hierarchical model and scene classification,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 563–579.
[6] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, “ORB: an efficient alternative to SIFT or SURF,” in IEEE International Conference on Computer Vision, ICCV, 2011, pp. 2564–2571.
[7] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, 2005, pp. 886– 893.
[8] R. Gomez-Ojeda, M. Lopez-Antequera, N. Petkov, and J. GonzalezJimenez, “Training a convolutional neural network for appearanceinvariant place recognition,” arXiv preprint arXiv:1505.07428, 2015.
[9] B. Sun and K. Saenko, “Deep coral: Correlation alignment for deep domain adaptation,” in European conference on computer vision. Springer, 2016, pp. 443–450.
[10] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” arXiv preprint arXiv:1503.02531, 2015.
[11] T. Hiroki and K. Tanaka, “Long-term knowledge distillation of visual place classifiers,” in 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019, pp. 541–546.
[12] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, “Planning and acting in partially observable stochastic domains,” Artificial intelligence, vol. 101, no. 1-2, pp. 99–134, 1998.
[13] E. Garcia-Fidalgo and A. Ortiz, “ibow-lcd: An appearance-based loopclosure detection approach using incremental bags of binary words,” IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 3051–3057, 2018.
[14] H. Morita, M. Hild, J. Miura, and Y. Shirai, “View-based localization in outdoor environments based on support vector learning,” in 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2005, pp. 2965–2970.
[15] 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. IEEE, 2012, pp. 1643–1649.
[16] M. Cummins and P. Newman, “Fab-map: Probabilistic localization and mapping in the space of appearance,” The International Journal of Robotics Research, vol. 27, no. 6, pp. 647–665, 2008.
[17] D. G. Lowe et al., “Object recognition from local scale-invariant features.” in iccv, vol. 99, no. 2, 1999, pp. 1150–1157.
[18] A. Oliva and A. Torralba, “Building the gist of a scene: The role of global image features in recognition,” vol. 155. Elsevier, 2006, pp. 23–36.
[19] D. Novotny, N. Ravi, B. Graham, N. Neverova, and A. Vedaldi, “C3dpo: Canonical 3d pose networks for non-rigid structure from motion,” in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 7688–7697.
[20] A. Walcott-Bryant, M. Kaess, H. Johannsson, and J. J. Leonard, “Dynamic pose graph slam: Long-term mapping in low dynamic environments,” in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 1871–1878.
[21] O. Erkent and I. Bozma, “Place representation in topological maps based on bubble space,” in 2012 IEEE International Conference on Robotics and Automation, May 2012, pp. 3497–3502.
[22] G. Kim, B. Park, and A. Kim, “1-day learning, 1-year localization: Longterm lidar localization using scan context image,” IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1948–1955, 2019.
[23] D. Fox, W. Burgard, and S. Thrun, “Active markov localization for mobile robots,” Robotics and Autonomous Systems, vol. 25, no. 3-4, pp. 195–207, 1998.
[24] S. K. Gottipati, K. Seo, D. Bhatt, V. Mai, K. Murthy, and L. Paull, “Deep active localization,” IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 4394–4401, 2019.
[25] L. Jiang, D. Meng, Q. Zhao, S. Shan, and A. G. Hauptmann, “Self-paced curriculum learning,” in Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
[26] A. Roberti, M. Carletti, F. Setti, U. Castellani, P. Fiorini, and M. Cristani, “Recognition self-awareness for active object recognition on depth images.” in BMVC, 2018, p. 15.
[27] J. Sock, S. Hamidreza Kasaei, L. Seabra Lopes, and T.-K. Kim, “Multiview 6d object pose estimation and camera motion planning using rgbd images,” in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 2228–2235.
[28] A. Doumanoglou, R. Kouskouridas, S. Malassiotis, and T.-K. Kim, “Recovering 6d object pose and predicting next-best-view in the crowd,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3583–3592.
[29] D. S. Chaplot, E. Parisotto, and R. Salakhutdinov, “Active neural localization,” arXiv preprint arXiv:1801.08214, 2018.
[30] A. M. Zhang and L. Kleeman, “Robust appearance based visual route following for navigation in large-scale outdoor environments,” The International Journal of Robotics Research, vol. 28, no. 3, pp. 331– 356, 2009.
[31] J. Straub, S. Hilsenbeck, G. Schroth, R. Huitl, A. Moller, and E. Stein- bach, “Fast relocalization for visual odometry using binary features,” in 2013 IEEE International Conference on Image Processing, 2013, pp. 2548–2552.
[32] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.
[33] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, “University of michigan north campus long-term vision and lidar dataset,” The International Journal of Robotics Research, pp. 1023–1035, 2015.
[34] H. Kurniawati, D. Hsu, and W. S. Lee, “SARSOP: efficient pointbased POMDP planning by approximating optimally reachable belief spaces,” in Robotics: Science and Systems IV, Eidgenossische Technische Hochschule Zurich, Zurich, Switzerland, June 25-28, 2008 , 2008.