On the Potential of Visual Place Recognition for Visual SLAM

Stefan Schubert
Chemnitz University of Technology
Email: stefan.schubert@etit.tu-chemnitz.de

Abstract—**This paper outlines the potential of visual place recognition (VPR), used for loop closure detection, to enhance camera-based SLAM. It highlights the gap between state-of-the-art VPR techniques proposed in the literature and those currently implemented in recent Visual SLAM systems. A preliminary experiment further demonstrates the potential benefit of integrating modern VPR methods into future SLAM pipelines.**

I. Introduction

Visual simultaneous localization and mapping (V-SLAM) is the key for truly autonomous robotic systems equipped with cameras operating in GNSS-denied or, more broadly, infrastructure-free environments such as (urban) canyons, caves, mines, extraterrestrial worlds, or disaster zones. V-SLAM is an active research field with numerous proposed methods and systems [1, 2, 3, 4]. An essential component of V-SLAM for building globally consistent maps in large-scale, long-term applications is visual place recognition (VPR) for loop closure detection. Like V-SLAM, VPR is an active research area [5, 6, 7, 8, 9]. However, despite the importance of VPR for V-SLAM, both research fields are quite independent, and recent V-SLAM systems rarely incorporate the latest advancements from the VPR literature. As a result, V-SLAM systems potentially miss the opportunity to achieve higher performance during large-scale, long-duration operation.

In this paper, I briefly outline the potential of modern VPR for enhanced V-SLAM. I begin with an overview of the diversity of VPR techniques, followed by a review of several recent V-SLAM systems to highlight the gap between existing and utilized VPR methods. Finally, I present a preliminary experiment demonstrating that V-SLAM performance improves when better-performing VPR methods are used.

II. Diversity of VPR Methods

There is a rich literature on VPR that proposes diverse methods across various categories to enhance performance and robustness. Below, some of the most important categories are presented to convey an impression of this diversity. A more detailed description of the following categories and corresponding techniques can be found in [11] as well as in the surveys [5, 6, 7, 8]. An introduction to the basics of VPR is provided in [9].

The main challenges for VPR during long-term operation are viewpoint changes, challenging conditions (e.g., fog, snow), and changing conditions (e.g., from day to night). To increase robustness against these challenges, a variety of *local and*

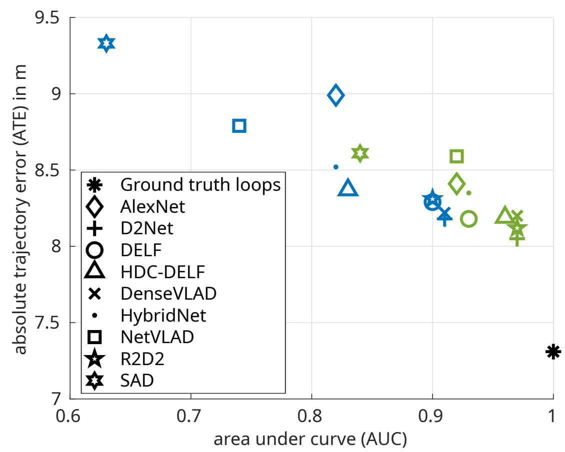

Fig. 1: Relation between performances of V-SLAM (ATE: lower is better) and its VPR component (AUC: higher is better) for all descriptors without and with postprocessing using SeqConv [10]

holistic image descriptors for image comparison have been developed, including CosPlace [12], D2Net [13], DELF [14], DenseVLAD [15], DinoV2SALAD [16], EigenPlaces [17], HDC-DELF [18], HybridNet [19], MixVPR [20], NetVLAD [21], R2D2 [22], and TransVPR [23]. Note that semantic information can also be integrated into descriptors [24, 25], e.g., to enable loop detection even when the viewpoint is in the opposite direction [24]. Alternatively, *image translation* methods have been proposed that convert the images of different conditions before feature extraction into a single reference condition using different learning-based techniques [26, 27, 28]. To improve the performance of any holistic image descriptor for image comparison, *descriptor transformations* based on principal component analysis [27, 29] or standardization [30, 31] have been applied. *Hierarchical VPR* combines holistic and local descriptors by first selecting matching candidates using the typically faster holistic descriptors, followed by a more accurate but slower verification using local descriptors [13, 32, 33]. Literature on the *efficient comparison of descriptors* focuses on selecting only a subset of image pairs for sparse descriptor comparison by leveraging additional knowledge about the dataset [34, 35]. Instead of computing holistic and local descriptors independently, *descriptor aggregation* combines a set of local descriptors from an image into a single holistic descriptor using approaches such as bag of visual words (BoW) [36, 37], vector of locally aggregated descriptors (VLAD) [15, 38], hyperdimensional computing (HDC) [18, 39], or deep

Visual SLAM / SfM System	Year	VPR System
AirSLAM [51]	2025	PLNet point features [51] + DBoW2 [37] + custom geometric consistency check with LightGlue [52]
Basalt [53]	2019	implicitly using ORB [54] and keypoint matching
DPV-SLAM++ [55]	2024	ORB [54] + DBoW2 [37] and proximity
DROID-SLAM [56]	2021	exhaustive computation of reprojection error between every frame combination
Gaussian Splatting SLAM [57]	2024	(no loop closure detection)
Kimera [58]	2020	ORB [54] + DBoW2 [37] + geometric verification
MASt3R-SLAM [59]	2024	MASt3R-encoder [60] + ASMK [61]
ORB-SLAM2 [62]	2017	ORB [54] + DBoW2 [37]
ORB-SLAM3 [63]	2021	ORB [54] + DBoW2 [37] with custom geometric and temporal consistency check
SuperVINS [64]	2025	SuperPoint [65] + DBoW3 [66]
VINS-Mono [67]	2018	Shi-Tomasi Corner Detector [68] + BRIEF [69] + DBoW2 [37]
COLMAP [70]	2016	RootSIFT [71] + scalable BoW [72]
GLOMAP [73]	2024	RootSIFT [71] + scalable BoW [72]
MASt3R-SfM [74]	2024	MASt3R-encoder [60] + ASMK [61]

TABLE I: Overview of V-SLAM (top) and SfM (bottom) systems with their publication year and used VPR system

learning [32]. *Multi-process fusion* can be used to combine the strengths of multiple descriptors (or potentially other VPR techniques) [18, 40, 41, 42]: For example, some descriptors may perform better in urban environments, while others may be more effective in natural environments [43] or in specific geographic regions such as Western cities [44]. For an even better recognition of known places, especially if the condition slightly changes between each loop, *place-specific descriptors or classifiers* can be used, e.g., based on experiences [45] or support vector machines [46]. Beyond image descriptors, the VPR performance can often be significantly improved using the well-studied *sequence-based methods* [10, 34, 47, 48, 49, 50], which aim to ensure temporal consistency when images are captured along a trajectory.

III. VPR Methods in V-SLAM Systems

Table I reviews several recent V-SLAM systems, along with related structure-from-motion (SfM) systems, including their publication year and the VPR system employed for loop closure detection or correspondence search, respectively. For an overview of six recent LiDAR-centric SLAM systems developed for operation in challenging environments during the SubT Challenge [75], please refer to [76].

Despite the diversity of VPR techniques, many state-of-the-art V-SLAM systems still rely on a limited set of older methods for loop closure detection. For instance, most systems in Table I use hand-crafted local descriptors such as ORB, RootSIFT, or Shi-Tomasi combined with BRIEF, even though the performance of their feature detectors is known to degrade under challenging or changing environmental conditions [15, 77, 78, 79]. Instead, descriptors from the VPR literature designed to handle severe viewpoint or environmental condition changes could be leveraged. However, some systems, such as AirSLAM, MASt3R-SLAM and SuperVINS, do employ deep-learned descriptors that are more robust to significant viewpoint and illumination changes. For computational efficiency, most V-SLAM systems actually adopt a hierarchical approach that first selects matching candidates using a holistic descriptor and then performs geometric verification. Typically, these holistic descriptors are constructed by aggregating local descriptors using methods such as DBoW2 from 2012 or ASMK from 2013. For a temporal consistency check, some of the V-SLAM systems use DBoW's integrated sequence method, which compares three consecutive (key)frames of a loop, although more sophisticated sequence-based methods exist in the VPR literature.

IV. Experimental Evaluation

The following preliminary experiment evaluates the potential of integrating modern VPR methods into a V-SLAM pipeline.

1) Setup: A lightweight pose-graph SLAM is implemented that combines odometry and loop closure detections using a factor graph [80] with Gaussian max mixture model [81] for robust optimization. For pairwise image comparison, one of six holistic or three local image descriptors is used. The resulting pairwise image similarities are optionally post-processed using one sequence-based method as a representative of the broad field of VPR methods beyond descriptors. For evaluation, five traversals through suburban streets from the St Lucia dataset [82] are used. Synthetic odometry data is extracted from GPS with 10% noise. All experiments are repeated 20 times.

2) Result: The results in Fig. 1 reveal a strong correlation between SLAM performance and VPR performance. This suggests that integrating more advanced VPR methods into future SLAM pipelines could further improve the overall SLAM performance. However, the result using ground truth loop closures also highlights the limit of VPR's impact, as SLAM accuracy also depends on (visual) odometry [83], the optimization backend, and the application and trajectory. For instance, V-SLAM on datasets with few, long loops is potentially more affected by missed loops than on datasets with frequent loops.

V. Conclusion

In this paper, I briefly outlined the potential of VPR for V-SLAM. The current gap between VPR techniques in the literature and those currently implemented in V-SLAM systems demonstrates that more sophisticated VPR methods could be integrated into loop closure detection. This could enable future V-SLAM systems to be more robust to viewpoint changes, challenging conditions and severe condition changes, which is particularly important during long-term applications in large-scale environments. A preliminary experiment demonstrated the correlation between VPR performance and V-SLAM performance, suggesting a benefit of using more advanced VPR techniques.

References

[1] Alejandro Fontan, Tobias Fischer, Javier Civera, and Michael Milford. VSLAM-LAB: A Comprehensive Framework for Visual SLAM Methods and Datasets. In *arXiv*, 2025. doi:10.48550/arXiv.2504.04457.

[2] Ali Rida Sahili, Saifeldin Hassan, Saber Muawiyah Sakhrieh, Jinane Mounsef, Noel Maalouf, Bilal Arain, and Tarek Taha. A survey of Visual SLAM methods. *IEEE Access*, 11:139643–139677, 2023. doi:10.1109/ACCESS.2023.3341489.

[3] Andréa Macario Barros, Maugan Michel, Yoann Moline, Gwenolé Corre, and Frédérick Carrel. A Comprehensive Survey of Visual SLAM Algorithms. *Robotics*, 11(1), 2022. doi:10.3390/robotics11010024.

[4] Ali Tourani, Hriday Bavle, Jose Luis Sanchez-Lopez, and Holger Voos. Visual SLAM: What are the current trends and what to expect? *Sensors*, 22(23), 2022. doi:10.3390/s22239297.

[5] Stephanie Lowry, Niko Sünderhauf, Paul Newman, John J. Leonard, David Cox, Peter Corke, and Michael J. Milford. Visual place recognition: A survey. *IEEE Transactions on Robotics (T-RO)*, 32(1):1–19, 2016. doi:10.1109/TRO.2015.2496823.

[6] Stefan Schubert and Peer Neubert. What makes visual place recognition easy or hard? In *arXiv*, 2021. doi:10.48550/arXiv.2106.12671.

[7] Sourav Garg, Tobias Fischer, and Michael Milford. Where is your place, visual place recognition? In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2021. doi:10.24963/ijcai.2021/603.

[8] Carlo Masone and Barbara Caputo. A survey on deep visual place recognition. *IEEE Access*, 9:19516–19547, 2021. doi:10.1109/ACCESS.2021.3054937.

[9] Stefan Schubert, Peer Neubert, Sourav Garg, Michael Milford, and Tobias Fischer. Visual Place Recognition: A Tutorial. *IEEE Robotics & Automation Magazine (RAM)*, 31(3):139–153, 2024. doi:10.1109/MRA.2023.3310859.

[10] Stefan Schubert, Peer Neubert, and Peter Protzel. Fast and memory efficient graph optimization via icm for visual place recognition. In *Robotics: Science and Systems (RSS)*, 2021. doi:10.15607/RSS.2021.XVII.091.

[11] Stefan Schubert. *Visual Place Recognition in Changing Environments using Additional Data-Inherent Knowledge*. PhD thesis, Chemnitz University of Technology, 2023. URL https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-872740.

[12] Gabriele Berton, Carlo Masone, and Barbara Caputo. Rethinking Visual Geo-localization for Large-Scale Applications. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. doi:10.1109/CVPR52688.2022.00483.

[13] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler. D2-Net: A trainable CNN for joint description and detection of local features. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. doi:10.1109/CVPR.2019.00828.

[14] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han. Large-scale image retrieval with attentive deep local features. In *International Conference on Computer Vision (ICCV)*, 2017. doi:10.1109/ICCV.2017.374.

[15] A. Torii, R. Arandjelović, J. Sivic, M. Okutomi, and T. Pajdla. 24/7 place recognition by view synthesis. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. doi:10.1109/CVPR.2015.7298790.

[16] Sergio Izquierdo and Javier Civera. Optimal Transport Aggregation for Visual Place Recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. doi:10.1109/CVPR52733.2024.01672.

[17] Gabriele Berton, Gabriele Trivigno, Barbara Caputo, and Carlo Masone. EigenPlaces: Training Viewpoint Robust Models for Visual Place Recognition. In *International Conference on Computer Vision (ICCV)*, 2023. doi:10.1109/ICCV51070.2023.01017.

[18] Peer Neubert and Stefan Schubert. Hyperdimensional computing as a framework for systematic aggregation of image descriptors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. doi:10.1109/CVPR46437.2021.01666.

[19] Z. Chen, A. Jacobson, N. Sünderhauf, B. Upcroft, L. Liu, C. Shen, I. Reid, and M. Milford. Deep learning features at scale for visual place recognition. In *International Conference on Robotics and Automation (ICRA)*, 2017. doi:10.1109/ICRA.2017.7989366.

[20] Amar Ali-Bey, Brahim Chaib-Draa, and Philippe Giguére. MixVPR: Feature Mixing for Visual Place Recognition. In *Winter Conference on Applications of Computer Vision (WACV)*, 2023. doi:10.1109/WACV56688.2023.00301.

[21] R. Arandjelović, P. Gronat, A. Torii, T. Pajdla, and J. Sivic. NetVLAD: CNN architecture for weakly supervised place recognition. *Trans. on Pattern Analysis and Machine Intelligence*, 40(6), 2018. doi:10.1109/TPAMI.2017.2711011.

[22] Jerome Revaud, Cesar De Souza, Martin Humenberger, and Philippe Weinzaepfel. R2D2: Reliable and repeatable detector and descriptor. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. doi:10.48550/arXiv.1906.06195.

[23] Ruotong Wang, Yanqing Shen, Weiliang Zuo, Sanping Zhou, and Nanning Zheng. TransVPR: Transformer-Based Place Recognition with Multi-Level Attention Aggregation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. doi:10.1109/CVPR52688.2022.01328.

[24] Sourav Garg, Niko Sünderhauf, and Michael Milford. Lost? appearance-invariant place recognition for opposite viewpoints using visual semantics. In *Robotics: Science and Systems (RSS)*, 2018. doi:10.15607/RSS.2018.XIV.022.

[25] Peer Neubert, Stefan Schubert, Kenny Schlegel, and Peter Protzel. Vector semantic representations as descriptors for visual place recognition. In *Robotics: Science and Systems (RSS)*, 2021. doi:10.15607/RSS.2021.XVII.083.

[26] P. Neubert, N. Sünderhauf, and P. Protzel. Appearance change prediction for long-term navigation across seasons. In *European Conference on Mobile Robots (ECMR)*, 2013. doi:10.1109/ECMR.2013.6698842.

[27] Stephanie Lowry and Michael J. Milford. Supervised and unsupervised linear learning techniques for visual place recognition in changing environments. *IEEE Transactions on Robotics (T-RO)*, 32(3):600–613, 2016. doi:10.1109/TRO.2016.2545711.

[28] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. Van Gool. Night-to-day image translation for retrieval-based localization. In *International Conference on Robotics and Automation (ICRA)*, 2019. doi:10.1109/ICRA.2019.8794387.

[29] Yang Liu and Hong Zhang. Visual loop closure detection with a compact image descriptor. In *International Conference on Intelligent Robots and Systems (IROS)*, 2012. doi:10.1109/IROS.2012.6386145.

[30] S. Garg, N. Sünderhauf, and M. Milford. Don't look back: Robustifying place categorization for viewpoint- and condition-invariant place recognition. In *International Conference on Robotics and Automation (ICRA)*, 2018. doi:10.1109/ICRA.2018.8461051.

[31] S. Schubert, P. Neubert, and P. Protzel. Unsupervised learning methods for visual place recognition in discretely and continuously changing environments. In *International Conference on Robotics and Automation (ICRA)*, 2020. doi:10.1109/ICRA40945.2020.9197044.

[32] Bingyi Cao, André Araujo, and Jack Sim. Unifying deep local and global features for image search. In *European Conference on Computer Vision (ECCV)*, 2020. doi:10.1007/978-3-030-58565-5_43.

[33] Fangming Yuan, Stefan Schubert, Peter Protzel, and Peer Neubert. Local positional graphs and attentive local features for a data and runtime-efficient hierarchical place recognition pipeline. *IEEE Robotics and Automation Letters (RA-L)*, 9(3):2686–2693, 2024. doi:10.1109/LRA.2024.3359552.

[34] Olga Vysotska and Cyrill Stachniss. Lazy data association for image sequences matching under substantial appearance changes. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):213–220, 2016. doi:10.1109/LRA.2015.2512936.

[35] S. Schubert, P. Neubert, and P. Protzel. Beyond ANN: Exploiting structural knowledge for efficient place recognition. In *International Conference on Robotics and Automation (ICRA)*, 2021. doi:10.1109/ICRA48506.2021.9561006.

[36] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *International Conference on Computer Vision (ICCV)*, 2003. doi:10.1109/ICCV.2003.1238663.

[37] D. Gálvez-López and J. D. Tardós. Bags of binary words for fast place recognition in image sequences. *IEEE Transactions on Robotics (T-RO)*, 28(5):1188–1197, 2012. doi:10.1109/TRO.2012.2197158.

[38] H. Jégou, M. Douze, C. Schmid, and P. Pérez. Aggregating local descriptors into a compact image representation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. doi:10.1109/CVPR.2010.5540039.

[39] P. Neubert, S. Schubert, and P. Protzel. An introduction to hyperdimensional computing for robotics. *KI - Künstliche Intelligenz, Special Issue: Reintegrating Artificial Intelligence and Robotics*, 33(4):319–330, 2019. doi:10.1007/s13218-019-00623-z.

[40] Stephen Hausler, Adam Jacobson, and Michael Milford. Multi-process fusion: Visual place recognition using multiple image processing methods. *IEEE Robotics and Automation Letters (RA-L)*, 4(2):1924–1931, 2019. doi:10.1109/LRA.2019.2898427.

[41] Hao Zhang, Fei Han, and Hua Wang. Robust multimodal sequence-based loop closure detection via structured sparsity. In *Robotics: Science and Systems (RSS)*, 2016. doi:10.15607/RSS.2016.XII.043.

[42] Maria Waheed, Michael Milford, Klaus McDonald-Maier, and Shoaib Ehsan. Improving visual place recognition performance by maximising complementarity. *IEEE Robotics and Automation Letters (RA-L)*, 6(3):5976–5983, 2021. doi:10.1109/LRA.2021.3088779.

[43] James Garforth and Barbara Webb. Lost in the woods? Place recognition for navigation in difficult forest environments. *Frontiers in Robotics and AI*, 7:1–9, 2020. doi:10.3389/frobt.2020.541770.

[44] Frederik Warburg, Søren Hauberg, Manuel López-Antequera, Pau Gargallo, Yubin Kuang, and Javier Civera. Mapillary street-level sequences: A dataset for lifelong place recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2623–2632, 2020. doi:10.1109/CVPR42600.2020.00270.

[45] Winston Churchill and Paul Newman. Experience-based navigation for long-term localisation. *The International Journal of Robotics Research (IJRR)*, 32(14):1645–1661, 2013. doi:10.1177/0278364913499193.

[46] Colin McManus, Ben Upcroft, and Paul Newmann. Scene Signatures: Localised and Point-less Features for Localisation. In *Robotics: Science and Systems (RSS)*, 2014. doi:10.15607/RSS.2014.X.023.

[47] M. J. Milford and G. F. Wyeth. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In *International Conference on Robotics and Automation (ICRA)*, 2012. doi:10.1109/ICRA.2012.6224623.

[48] Tayyab Naseer, Wolfram Burgard, and Cyrill Stachniss. Robust visual localization across seasons. *IEEE Transactions on Robotics (T-RO)*, 34(2):289–302, 2018. doi:10.1109/TRO.2017.2788045.

[49] S. Garg, B. Harwood, G. Anand, and M. Milford. Delta descriptors: Change-based place representation for robust visual localization. *IEEE Robotics and Automation Letters (RA-L)*, 5(4):5120–5127, 2020. doi:10.1109/LRA.2020.3005627.

[50] Peer Neubert, Stefan Schubert, and Peter Protzel. A neurologically inspired sequence processing model for

mobile robot place recognition. *IEEE Robotics and Automation Letters (RA-L)*, 4(4):3200–3207, 2019. doi:10.1109/LRA.2019.2927096.
[51] Kuan Xu, Yuefan Hao, Shenghai Yuan, Chen Wang, and Lihua Xie. AirSLAM: An Efficient and Illumination-Robust Point-Line Visual SLAM System. *IEEE Transactions on Robotics (T-RO)*, 41:1673–1692, 2025. doi:10.1109/TRO.2025.3539171.
[52] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue: Local Feature Matching at Light Speed. In *International Conference on Computer Vision (ICCV)*, 2023. doi:10.1109/ICCV51070.2023.01616.
[53] V. Usenko, N. Demmel, D. Schubert, J. Stueckler, and D. Cremers. Visual-inertial mapping with non-linear factor recovery. *IEEE Robotics and Automation Letters (RA-L)*, 5(2):422–429, 2020. doi:10.1109/LRA.2019.2961227.
[54] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *International Conference on Computer Vision (ICCV)*, 2011. doi:10.1109/ICCV.2011.6126544.
[55] Lahav Lipson, Zachary Teed, and Jia Deng. Deep Patch Visual SLAM. In *European Conference on Computer Vision (ECCV)*, 2024. doi:10.1007/978-3-031-72627-9_24.
[56] Zachary Teed and Jia Deng. DROID-SLAM: Deep Visual SLAM for Monocular, Stereo, and RGB-D Cameras. *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. doi:10.48550/arXiv.2108.10869.
[57] Hidenobu Matsuki, Riku Murai, Paul H. J. Kelly, and Andrew J. Davison. Gaussian Splatting SLAM. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. doi:10.1109/CVPR52733.2024.01708.
[58] Antoni Rosinol, Marcus Abate, Yun Chang, and Luca Carlone. Kimera: an open-source library for real-time metric-semantic localization and mapping. In *International Conference on Robotics and Automation (ICRA)*, 2020. doi:10.1109/ICRA40945.2020.9196885.
[59] Riku Murai, Eric Dexheimer, and Andrew J. Davison. MASt3R-SLAM: Real-Time Dense SLAM with 3D Reconstruction Priors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. doi:10.48550/arXiv.2412.12392.
[60] Vincent Leroy, Yohann Cabon, and Jerome Revaud. Grounding image matching in 3d with mast3r. In *European Conference on Computer Vision (ECCV)*, 2025. doi:10.1007/978-3-031-73220-1_5.
[61] Giorgos Tolias, Yannis Avrithis, and Hervé Jégou. To aggregate or not to aggregate: Selective match kernels for image search. In *International Conference on Computer Vision (ICCV)*, 2013. doi:10.1109/ICCV.2013.177.
[62] Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics (T-RO)*, 33(5):1255–1262, 2017. doi:10.1109/TRO.2017.2705103.
[63] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. *IEEE Transactions on Robotics (T-RO)*, 37(6):1874–1890, 2021. doi:10.1109/TRO.2021.3075644.
[64] Hongkun Luo, Yang Liu, Chi Guo, Zengke Li, and Weiwei Song. SuperVINS: A Real-Time Visual-Inertial SLAM Framework for Challenging Imaging Conditions. *IEEE Sensors Journal*, 2025. doi:10.1109/JSEN.2025.3556257.
[65] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperPoint: Self-Supervised Interest Point Detection and Description. In *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. doi:10.1109/CVPRW.2018.00060.
[66] https://github.com/rmsalinas/DBow3, 2017.
[67] Tong Qin, Peiliang Li, and Shaojie Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics (T-RO)*, 34 (4):1004–1020, 2018. doi:10.1109/TRO.2018.2853729.
[68] Jianbo Shi and Tomasi. Good features to track. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994. doi:10.1109/CVPR.1994.323794.
[69] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF: Binary Robust Independent Elementary Features. In *European Conference on Computer Vision (ECCV)*, 2010. doi:10.1007/978-3-642-15561-1_56.
[70] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. doi:10.1109/CVPR.2016.445.
[71] Relja Arandjelović and Andrew Zisserman. Three things everyone should know to improve object retrieval. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. doi:10.1109/CVPR.2012.6248018.
[72] Johannes L. Schönberger, True Price, Torsten Sattler, Jan-Michael Frahm, and Marc Pollefeys. A vote-and-verify strategy for fast spatial verification in image retrieval. In *Asian Conference on Computer Vision (ACCV)*, 2017. doi:10.1007/978-3-319-54181-5_21.
[73] Linfei Pan, Daniel Barath, Marc Pollefeys, and Johannes Lutz Schönberger. Global Structure-from-Motion Revisited. In *European Conference on Computer Vision (ECCV)*, 2024. doi:10.1007/978-3-031-73661-2_4.
[74] Bardienus Duisterhof, Lojze Zust, Philippe Weinzaepfel, Vincent Leroy, Yohann Cabon, and Jerome Revaud. MASt3R-SfM: a Fully-Integrated Solution for Unconstrained Structure-from-Motion. In *arXiv*, 2024. doi:10.48550/arXiv.2409.19152.
[75] Shibo Zhao, Yuanjun Gao, Tianhao Wu, Damanpreet Singh, Rushan Jiang, Haoxiang Sun, Mansi Sarawata, Yuheng Qiu, Warren Whittaker, Ian Higgins, Yi Du, Shaoshu Su, Can Xu, John Keller, Jay Karhade, Lucas Nogueira, Sourojit Saha, Ji Zhang, Wenshan Wang, Chen Wang, and Sebastian Scherer. SubT-MRS Dataset: Pushing SLAM Towards All-weather Environments. In *Conference on Computer Vision and Pattern Recognition (CVPR)*,

2024. doi:10.1109/CVPR52733.2024.02137.

[76] Kamak Ebadi, Lukas Bernreiter, Harel Biggie, Gavin Catt, Yun Chang, Arghya Chatterjee, Christopher E. Denniston, Simon-Pierre Deschênes, Kyle Harlow, Shehryar Khattak, Lucas Nogueira, Matteo Palieri, Pavel Petráček, Matěj Petrlík, Andrzej Reinke, Vít Krátký, Shibo Zhao, Ali-akbar Agha-mohammadi, Kostas Alexis, Christoffer Heckman, Kasra Khosoussi, Navinda Kottege, Benjamin Morrell, Marco Hutter, Fred Pauling, François Pomerleau, Martin Saska, Sebastian Scherer, Roland Siegwart, Jason L. Williams, and Luca Carlone. Present and Future of SLAM in Extreme Environments: The DARPA SubT Challenge. *IEEE Transactions on Robotics (T-RO)*, 40:936–959, 2024. doi:10.1109/TRO.2023.3323938.

[77] Christoffer Valgren and Achim Lilienthal. SIFT, SURF and Seasons: Long-term Outdoor Localization Using Local Features. In *European Conference on Mobile Robots (ECMR)*, 2007. URL http://ecmr07.informatik.uni-freiburg.de/proceedings/ECMR07_0050.pdf.

[78] Paul Furgale and Timothy D. Barfoot. Visual teach and repeat for long-range rover autonomy. *Journal of Field Robotics*, 27(5):534–560, 2010. doi:https://doi.org/10.1002/rob.20342.

[79] Peer Neubert. *Superpixels and their Application for Visual Place Recognition in Changing Environments*. PhD thesis, Chemnitz University of Technology, 2015. URL https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa-190241.

[80] Frank Dellaert and Michael Kaess. *Factor Graphs for Robot Perception*, volume 6. Foundations and Trends in Robotics, 2017. doi:10.1561/2300000043.

[81] Edwin Olson and Pratik Agarwal. Inference on networks of mixtures for robust robot mapping. *The International Journal of Robotics Research (IJRR)*, 32(7):826–840, 2013. doi:10.1177/0278364913479413.

[82] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth. FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day. In *International Conference on Robotics and Automation (ICRA)*, 2010. doi:10.1109/ROBOT.2010.5509547.

[83] Davide Scaramuzza and Friedrich Fraundorfer. Visual Odometry [Tutorial]. *IEEE Robotics & Automation Magazine (RAM)*, 18(4):80–92, 2011. doi:10.1109/MRA.2011.943233.