



# New Trend in Front-End Techniques of Visual SLAM: From Hand-Engineered Features to Deep-Learned Features

Yue Wang<sup>1,2(✉)</sup>, Yu Fu<sup>2</sup>, Ruixue Zheng<sup>2</sup>, Le Wang<sup>3</sup>, and Jianzhong Qi<sup>3</sup>

<sup>1</sup> Key Laboratory of Electronics and Information Technology for Space Systems, Chinese Academy of Sciences, Beijing 100190, China

wangyue@nssc.ac.cn

<sup>2</sup> Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China

<sup>3</sup> Department of Electronic and Information Engineering, North China University of Technology, Beijing 100144, China

**Abstract.** Visual simultaneous localization and mapping (V-SLAM) technique plays a key role in perception of autonomous mobile robots, augmented/mixed/virtual reality, as well as spatial AI applications. This paper gives a very concise survey about the front-end module of a V-SLAM system, which is charge of feature extraction, short/long-term data association with outlier rejection, as well as variable initialization. Visual features are mainly salient and repeatable points, called keypoints, their traditional extractors and matchers are hand-engineered and not robust to viewpoint/illumination/seasonal change, which is crucial for long-term autonomy of mobile robots. Therefore, new trend about deep-learning-based keypoint extractors and matchers are introduced to enhance the robustness of V-SLAM systems even under challenging conditions.

**Keywords:** 3D vision · Data association · Deep learning · Feature extraction · Keypoints · Simultaneous localization and mapping (SLAM) · Visual odometry (VO)

## 1 Introduction

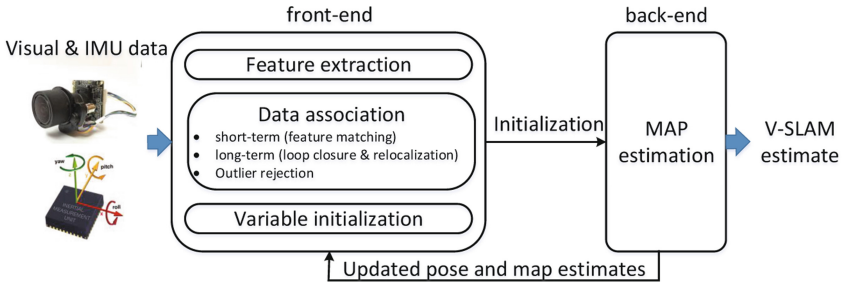
The visual simultaneous localization and mapping (V-SLAM) system adopts cameras (usually with inertial measurement units (IMUs)) to simultaneously inference its own state (e.g. pose) and build a consistent surrounding map [1]. V-SLAM systems are widely used by autonomous mobile robots, especially the

---

Y. Wang—This work was supported by the Pre-Research Project of Space Science (No. XDA15014700), the National Natural Science Foundation of China (No. 61601328), the Scientific Research Plan Project of the Committee of Education in Tianjin (No. JW1708), and the Doctor Foundation of Tianjin Normal University (No. 52XB1417), and the Beijing Natural Science Foundation (No. 4204095).

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022  
Q. Liang et al. (Eds.): Artificial Intelligence in China, LNEE 854, pp. 298–307, 2022.  
[https://doi.org/10.1007/978-981-16-9423-3\\_38](https://doi.org/10.1007/978-981-16-9423-3_38)

resource-limited agile micro drones [2], augmented/mixed/virtual reality devices [3], as well as spatial AI applications [4, 5]. Recently, with the huge tide of deep learning (DL) techniques [6, 7], some end-to-end DL-based V-SLAM or visual odometry (VO) systems are proposed [8–11]. In this paper, we focus on the traditional de-facto framework of V-SLAM systems with standard cameras, as shown in Fig. 1, which contains two modules: a front-end and a back-end. The front-end is responsible for processing visual and IMU data, including feature extraction, short/long-term data association with outlier rejection, as well as initial estimation of the current pose and newly detected landmark positions. The back-end takes the initialization information from the front-end, and makes a maximum-a-posteriori (MAP) estimation based on a factor graph to refine both poses and landmark positions [12]. This paper focuses on the new trend in front-end techniques of V-SLAM systems.



**Fig. 1.** The de-facto framework of a V-SLAM system.

## 2 Hand-Engineered Features

Traditionally, visual features are mainly restricted to salient and repeatable points (called keypoints), due to the unreliable extraction of high-level geometric features (e.g. lines or edges) using unlearning methods. Besides keypoints, dense methods using all pixel information (e.g. optical flow [13] or correspondence-free method [14]) have also been adopted to inference ego-motion with small motion assumption. In the early research of stereo VO for Mars rovers [15–18], keypoints were tracked among successive images from nearby viewpoints. Nister proposed that keypoints could be independently extracted in all images and then matched in his landmark paper [19]. This method then became the dominant approach because it can successfully work with a large motion/viewpoint change [20].

Keypoint-based feature extraction includes two stages: keypoint detection and keypoint description. The keypoint detectors can be divided into two categories: corner detectors and blob detectors [21]. Traditional keypoint extractors are hand-engineered: corner detectors include Moravec [22], Forstner [23], Harris [24], Shi-Tomasi [25], and FAST/FASTER [26, 27]; blob detectors include SIFT [28], SURF [29], SENSURE [30], RootSIFT [31], and KAZE [32, 33]; keypoint

descriptors include SSD/NCC [34], census transform [35], SIFT [28], GLOH [36], SURF [29], DAISY [37], BRIEF [38], ORB [39], BRISK [40], and KAZE [32, 33].

Data association (also called feature matching) is commonly conducted by comparing similarity measurements between keypoint descriptors along with a mutual consistency check procedure [21]. With a prior knowledge of motion constraints (e.g. from IMU sensors or constant velocity assumptions) or stereo's epipolar line constraints, the time used for data association can be shortened by restricting the searching space [17, 41, 42]. Due to the visual aliasing, wrong data associations (called outliers) are unavoidable to both short-term feature matching and long-term loop closure. Therefore, consensus set search (e.g. RANSAC) [43–48] and geometric information (i.e. pose and map estimations) from the back-end are usually adopted to remove outliers.

Since traditional back-end of V-SLAM systems relies on local iterative optimization algorithms (e.g. Gauss-Newton) [12], a fairly good initialization of variables (e.g. 6-DoF poses and 3D coordinate of keypoints) is required. The 6-DoF poses can be estimated via multiview geometry knowledge [18, 34, 49] using keypoints. 3D coordinate of keypoints are obtained by triangulation with careful keyframe selection [50, 51].

### 3 Deep-Learned Features

Traditional hand-engineered keypoint extractors and matchers are not robust to viewpoint/illumination/seasonal change, which is crucial for long-term autonomy of mobile robots. While deep learning, especially the convolutional neural network (CNN) [52], is good at feature extraction and processing. Therefore, new trend about DL-based keypoint extractors and matchers are proposed to enhance the robustness of V-SLAM systems even under challenging conditions.

The DL-based keypoint extractors can be divided into three categories: detect-then-describe, jointly detect-and-describe, describe-to-detect. For the detect-then-describe method, keypoint detectors include TILDE [53], covariant feature detector [54], TCDet [55], MagicPoint [56], Quad-Networks [57], texture feature detector [58], and Key.Net [59]; keypoint descriptors include convex optimized descriptor [60], Deepdesc [61], TFeat [62], UCN [63], L2-Net [64], HardNet [65], averaging precision ranking [66], GeoDesc [67], RDRL [68], Log-PolarDesc [69], ContextDesc [70], SOSNet [71], GIFT [72], and CAPS descriptor [73]. Jointly detect-and-describe methods include LIFT [74], DELE [75], LF-Net [76], SuperPoint [77], UnsuperPoint [78], GCNv2 [79], D2-Net [80], RF-Net [81], R2D2 [82], ASLFeat [83], and UR2KiD [84]. Different from the above two categories, Tian proposed a describe-to-detect (D2D) method [85], which selects keypoints based on the dense descriptors information.

After feature extracting, robust keypoint matching methods include content networks (CNe) [86], deep fundamental matrix estimation [87], NG-RANSAC [88], NM-Net [89], Order-Aware Network [90], ACNe [91], and SuperGlue [92].

Besides above extract-then-match methods, some end-to-end keypoint extract-and-match methods have also been proposed, they include NCNet [93], KP2D [94], Sparse-NCNet [95], LoFTR [96], and Patch2Pix [97].

Some datasets and benchmarks are released to train and evaluate different DL networks, such as Brown and Lowe’s dataset [98], HPatches [99], matching in the dark (MID) dataset [100], image matching challenge [101], and long-term visual localization benchmark [102, 103].

## 4 Conclusions

In this paper, we give a very concise survey of the front-end techniques of V-SLAM systems, emphasizing the DL-based keypoint extracting and matching techniques. Future research directions include extraction and association of high-level geometric features (e.g. lines [104, 105], symmetry [106], and holistic 3D structures [107]) as well as semantic object-level features [108–110].

## References

1. Cadena, C., et al.: Past, present, and future of simultaneous localization and mapping: toward the robust-perception age. *IEEE Trans. Robot.* **32**(6), 1309–1332 (2016)
2. Vidal, A.R., Rebecq, H., Horstschaefer, T., Scaramuzza, D.: Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios. *IEEE Robot. Autom. Lett.* **3**(2), 994–1001 (2018)
3. Microsoft HoloLens 2. <http://www.microsoft.com/en-us/hololens>
4. Davison, A.J.: FutureMapping: the computational structure of spatial AI systems. ArXiv Preprint [arXiv:1803.11288](https://arxiv.org/abs/1803.11288) (2018)
5. Davison, A.J., Ortiz, J.: FutureMapping 2: Gaussian belief propagation for spatial AI. ArXiv Preprint [arXiv:1910.14139](https://arxiv.org/abs/1910.14139) (2019)
6. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436–444 (2015)
7. Bengio, Y., LeCun, Y., Hinton, G.: Deep learning for AI. *Commun. ACM* **64**(7), 58–65 (2021)
8. Kendall, A., Grimes, M., Cipolla, R.: PoseNet: a convolutional network for real-time 6-DOF camera relocalization. In: *IEEE 12th International Conference on Computer Vision (ICCV)*, 7–13 December 2015, Santiago, Chile, pp. 1–9 (2015)
9. Wang, S., Clark, R., Wen, H., Trigoni, N.: DeepVO: towards end-to-end visual odometry with deep recurrent convolutional neural networks. In: *IEEE International Conference on Robotics and Automation (ICRA)*, 29 May–3 June 2017, Singapore, pp. 1–8 (2017)
10. Yang, N., Stumberg, L.V., Wang, R., Cremers, D.: D3VO: deep depth, deep pose and deep uncertainty for monocular visual odometry. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13–19 June 2020, Seattle, WA, USA, pp. 1–12 (2020)
11. Chen, C., Wang, B., Lu, C.X., Trigoni, N., Markham, A.: A survey on deep learning for localization and mapping: towards the age of spatial machine intelligence. ArXiv Preprint [arXiv:2006.12567v2](https://arxiv.org/abs/2006.12567v2) (2020)
12. Wang, Y., Peng, X.: New trend in back-end techniques of visual SLAM: from local iterative solvers to robust global optimization. In: *International Conference on Artificial Intelligence in China (CHINAAI)*, 21–22 August 2021, pp. 1–10 (2021)
13. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(3), 611–625 (2018)

14. Makadia, A., Geyer, C., Daniilidis, K.: Correspondence-free structure from motion. *Int. J. Comput. Vis.* **75**(3), 311–327 (2007)
15. Moravec, H.: Obstacle avoidance and navigation in the real world by a seeing robot rover. Ph.D. dissertation, Stanford University, USA (1980)
16. Cheng, Y., Maimone, M.W., Matthies, L.: Visual odometry on the mars exploration rovers. *IEEE Robot. Automat. Mag.* **13**(2), 54–62 (2006)
17. Maimone, M.W., Cheng, Y., Matthies, L.: Two years of visual odometry on the mars exploration rovers: field reports. *J. Field Robot.* **24**(3), 169–186 (2007)
18. Scaramuzza, D., Fraundorfer, F.: Visual odometry: part I: the first 30 years and fundamentals. *IEEE Robot. Autom. Mag.* **18**(4), 80–92 (2011)
19. Nister, D., Naroditsky, O., Bergen, J.: Visual odometry. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27 June–2 July 2004, Washington, DC, USA, pp. 1–8 (2004)
20. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras. *IEEE Trans. Robot.* **33**(5), 1255–1262 (2017)
21. Fraundorfer, F., Scaramuzza, D.: Visual odometry: part II: matching, robustness, optimization, and applications. *IEEE Robot. Autom. Mag.* **19**(2), 78–90 (2012)
22. Moravec, H.: Towards automatic visual obstacle avoidance. In: *5th International Joint Conference on Artificial Intelligence*, p. 584, August 1977
23. Forstner, W.: A feature based correspondence algorithm for image matching. *Int. Arch. Photogram. Remote Sens.* **26**(3), 150–166 (1986)
24. Harris, C., Pike, J.: 3D positional integration from image sequences. In: *3rd Alvey Vision Conference*, Cambridge, pp. 233–236, September 1987
25. Tomasi, C., Shi, J.: Good features to track. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21–23 June 1994, Seattle, WA, USA, pp. 1–8 (1994)
26. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006). [https://doi.org/10.1007/11744023\\_34](https://doi.org/10.1007/11744023_34)
27. Rosten, E., Porter, R., Drummond, T.: Faster and better: a machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(1), 105–119 (2010)
28. Lowe, D.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **20**(2), 91–110 (2003)
29. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006*. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). [https://doi.org/10.1007/11744023\\_32](https://doi.org/10.1007/11744023_32)
30. Agrawal, M., Konolige, K., Blas, M.R.: CenSurE: center surround extremas for realtime feature detection and matching. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008*. LNCS, vol. 5305, pp. 102–115. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-88693-8\\_8](https://doi.org/10.1007/978-3-540-88693-8_8)
31. Arandjelovic, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 16–21 June 2012, Providence, RI, USA, pp. 1–8 (2012)
32. Alcantarilla, P.F., Bartoli, A., Davison, A.J.: KAZE features. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012*. LNCS, vol. 7577, pp. 214–227. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-33783-3\\_16](https://doi.org/10.1007/978-3-642-33783-3_16)

33. Alcantarilla, P.F., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: British Machine Vision Conference (BMVC), 9–13 September 2013, Bristol, UK, pp. 1–11 (2013)
34. Ma, Y., Soatto, S., Kosčák, J., Sastry, S.S.: An Invitation to 3-D Vision: From Images to Geometric Models. Springer, New York (2004). <https://doi.org/10.1007/978-0-387-21779-6>
35. Zabih, R., Woodfill, J.: Non-parametric local transforms for computing visual correspondence. In: Eklundh, J.-O. (ed.) ECCV 1994. LNCS, vol. 801, pp. 151–158. Springer, Heidelberg (1994). <https://doi.org/10.1007/BFb0028345>
36. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
37. Tola, E., Lepetit, V., Fua, P.: DAISY: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 815–830 (2010)
38. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15561-1\\_56](https://doi.org/10.1007/978-3-642-15561-1_56)
39. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: IEEE International Conference on Computer Vision (ICCV), 6–13 November 2011, Barcelona, Spain, pp. 2564–2571 (2011)
40. Leutenegger, S., Chli, M., Siegwart, R.: BRISK: binary robust invariant scalable keypoints. In: IEEE International Conference on Computer Vision (ICCV), 6–13 November 2011, Barcelona, Spain, pp. 2548–2555 (2011)
41. Davison, A.: Real-time simultaneous localization and mapping with a single camera. In: IEEE International Conference on Computer Vision (ICCV), 14–17 October 2003, Nice, France, pp. 1403–1410 (2003)
42. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**(1–3), 7–42 (2002)
43. Fishler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)
44. Latif, Y., Cadena, C., Neira, J.: Robust loop closing over time for pose graph SLAM. Intl. J. Robot. Res. **32**(14), 1611–1626 (2013)
45. Mangelson, J.G., Dominic, D., Eustice, R.M., Vasudevan, R.: Pairwise consistent measurement set maximization for robust multi-robot map merging. In: IEEE International Conference on Robotics and Automation (ICRA), 21–25 May 2018, Brisbane, Australia, pp. 1–8 (2018)
46. Yang, J., Huang, Z., Quan, S., Qi, Z., Zhang, Y.: SAC-COT: sample consensus by sampling compatibility triangles in graphs for 3-D point cloud registration. IEEE Trans. Geosci. Remote Sens. (Early Access)
47. Yang, H., Antonante, P., Tzoumas, V., Carlone, L.: Graduated non-convexity for robust spatial perception: from non-minimal solvers to global outlier rejection. IEEE Robot. Autom. Lett. **5**(2), 1127–1134 (2020). (Best Paper Award Honorable Mention)
48. Shi, J., Yang, H., Carlone, L.: ROBIN: a graph-theoretic approach to reject outliers in robust estimation using invariants. In: IEEE International Conference on Robotics and Automation (ICRA), 30 May–5 June 2021, Xi'an, China, pp. 1–16 (2021)
49. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, UK (2003)

50. Mourikis, A.I., Roulletiotis, S.I.: A multi-state constraint Kalman filter for vision-aided inertial navigation. In: IEEE International Conference on Robotics and Automation (ICRA), 10–14 April 2007, Roma, Italy, pp. 1–8 (2007)
51. Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., Furgale, P.: Keyframe-based visual-inertial odometry using nonlinear optimization. *Intl. J. Robot. Res.* **34**(3), 314–334 (2015)
52. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
53. Verdie, Y., Yi, K.M., Fua, P., Lepetit, V.: TILDE: a temporally invariant learned DEtector. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7–12 June 2015, Boston, MA, USA, pp. 1–10 (2015)
54. Lenc, K., Vedaldi, A.: Learning covariant feature detectors. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9915, pp. 100–117. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-49409-8\\_11](https://doi.org/10.1007/978-3-319-49409-8_11)
55. Zhang, X., Yu, F.X., Karaman, S., Chang, S.-F.: Learning discriminative and transformation covariant local feature detectors, In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017, Honolulu, HI, USA, pp. 1–9 (2017)
56. DeTone, D., Malisiewicz, T., Rabinovich, A.: Toward geometric deep SLAM. ArXiv Preprint [arXiv:1707.07410v1](https://arxiv.org/abs/1707.07410v1) (2017)
57. Savinov, N., Seki, A., Ladický, L., Sattler, T., Pollefeys, M.: Quad-networks: unsupervised learning to rank for interest point detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017, Honolulu, HI, USA, pp. 1–9 (2017)
58. Zhang, L., Rusinkiewicz, S.: Learning to detect features in texture images. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18–22 June 2018, Salt Lake City, UT, USA, pp. 1–9 (2018)
59. Laguna, A.B., Riba, E., Ponsa, D., Mikolajczyk, K.: Key.Net: keypoint detection by handcrafted and learned CNN filters. In: IEEE/CVF International Conference on Computer Vision (ICCV), 27 October–2 November 2019, Seoul, Korea (South), pp. 1–9 (2019)
60. Simonyan, K., Vedaldi, A., Zisserman, A.: Learning local feature descriptors using convex optimisation. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(8), 1573–1585 (2014)
61. Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F.: Discriminative learning of deep convolutional feature point descriptors. In: IEEE International Conference on Computer Vision (ICCV), 7–13 December 2015, Santiago, Chile, pp. 1–9 (2015)
62. Balntas, V., Riba, E., Ponsa, D., Mikolajczyk, K.: Learning local feature descriptors with triplets and shallow convolutional neural networks. In: British Machine Vision Conference (BMVC), 19–22 September 2016, York, UK, pp. 1–11 (2016)
63. Choy, C.B., Gwak, J., Savarese, S., Chandraker, M.: Universal correspondence network. In: Conference on Neural Information Processing Systems (NeurIPS), 5–10 December 2016, Barcelona, Spain, pp. 1–9 (2016)
64. Tian, Y., Fan, B., Wu, F.: L2-net: deep learning of discriminative patch descriptor in euclidean space. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017, Honolulu, HI, USA, pp. 1–9 (2017)
65. Mishchuk, A., Mishkin, D., Radenović, F., Matas, J.: Working hard to know your neighbor’s margins: local descriptor learning loss. In: Conference on Neural Information Processing Systems (NeurIPS), 4–9 December 2017, Long Beach, NY, USA, pp. 1–12 (2017)



66. He, K., Lu, Y., Sclaroff, S.: Local descriptors optimized for averaging precision. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18–22 June 2018, Salt Lake City, UT, USA, pp. 1–10 (2018)
67. Luo, Z., et al.: GeoDesc: learning local descriptors by integrating geometry constraints. In: European Conference on Computer Vision (ECCV), 8–14 September 2018, Munich, Germany, pp. 170–185 (2018)
68. Yu, X., et al.: Unsupervised extraction of local image descriptors via relative distance ranking loss. In: IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 27 October–2 November 2019, Seoul, Korea (South), pp. 1–10 (2019)
69. Ebel, P., Trulls, E., Yi, K.M., Fua, P., Mishchuk, A.: Beyond cartesian representations for local descriptors. In: IEEE/CVF International Conference on Computer Vision (ICCV), 27 October–2 November 2019, Seoul, Korea (South), pp. 1–10 (2019)
70. Luo, Z., et al.: ContextDesc: local descriptor augmentation with cross-modality context. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16–20 June 2019, Long Beach, CA, USA, pp. 1–10 (2019)
71. Tian, Y., Yu, X., Fan, B., Wu, F., Heijnen, H., Balntas, V.: SOSNet: second order similarity regularization for local descriptor learning. 16–20 June 2019, Long Beach, CA, USA, pp. 1–10 (2019)
72. Liu, Y., Shen, Z., Lin, Z., Peng, S., Bao, H., Zhou, X.: GIFT: learning transformation-invariant dense visual descriptors via group CNNs. In: Conference on Neural Information Processing Systems (NeurIPS), 8–14 December 2019, Vancouver, Canada, pp. 1–12 (2019)
73. Wang, Q., Zhou, X., Hariharan, B., Snavely, N.: Learning feature descriptors using camera pose supervision. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12346, pp. 757–774. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58452-8\\_44](https://doi.org/10.1007/978-3-030-58452-8_44)
74. Yi, K.M., Trulls, E., Lepetit, V., Fua, P.: LIFT: learned invariant feature transform. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 467–483. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46466-4\\_28](https://doi.org/10.1007/978-3-319-46466-4_28)
75. Noh, H., Araujo, A., Sim, J., Weyand, T., Han, B.: Large-scale image retrieval with attentive deep local features. In: IEEE International Conference on Computer Vision (ICCV), 22–29 October 2017, Venice, Italy, pp. 1–10 (2017)
76. Ono, Y., Trulls, E., Fua, P., Yi, K.M.: LF-net: learning local features from images. In: Conference on Neural Information Processing Systems (NeurIPS), 2–8 December 2018, Montreal, Canada, pp. 1–11 (2018)
77. DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperPoint: self-supervised interest point detection and description. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 18–22 June 2018, Salt Lake City, UT, USA, pp. 1–13 (2018)
78. Christiansen, P.H., Kragh, M.F., Brodskiy, Y., Karstoft, H.: UnsuperPoint: end-to-end unsupervised interest point detector and descriptor. ArXiv Preprint [arXiv:1907.04011](https://arxiv.org/abs/1907.04011) (2019)
79. Tang, J., Ericson, L., Folkesson, J., Jensfelt, P.: GCNv2: efficient correspondence prediction for real-time SLAM. *IEEE Robot. Autom. Lett.* **4**(4), 3505–3512 (2019)
80. Dusmanu, M., et al.: D2-Net: a trainable cnn for joint description and detection of local features. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16–20 June 2019, Long Beach, CA, USA, pp. 1–10 (2019)



81. Shen, X., et al.: RF-net: an end-to-end image matching network based on receptive field. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16–20 June 2019, Long Beach, CA, USA, pp. 1–9 (2019)
82. Revaud, J., Weinzaepfel, P., Souza, C.D., Humenberger, M.: R2D2: repeatable and reliable detector and descriptor. In: Conference on Neural Information Processing Systems (NeurIPS), 8–14 December 2019, Vancouver, Canada, pp. 1–11 (2019)
83. Luo, Z., et al.: ASLFeat: learning local features of accurate shape and localization. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 14–19 June 2020, pp. 1–10 (2020)
84. Yang, T.-Y., Nguyen, D.-K., Heijnen, H., Balntas, V.: UR2KiD: unifying retrieval, keypoint detection, and keypoint description without local correspondence supervision. ArXiv Preprint [arXiv:2001.07252](https://arxiv.org/abs/2001.07252) (2020)
85. Tian, Y., Balntas, V., Ng, T., Barroso-Laguna, A., Demiris, Y., Mikolajczyk, K.: D2D: keypoint extraction with describe to detect approach. In: Asian Conference on Computer Vision (ACCV), 30 November–4 December 2020, pp. 1–18 (2020)
86. Yi, K.M., Trulls, E., Ono, Y., Lepetit, V., Salzmann, M., Fua, P.: Learning to find good correspondences. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18–22 June 2018, Salt Lake City, UT, USA, pp. 1–9 (2018)
87. Ranftl, R., Koltun, V.: Deep fundamental matrix estimation. In: European Conference on Computer Vision (ECCV), 8–14 September 2018, Munich, Germany, pp. 292–309 (2018)
88. Brachmann, E., Rother, C.: Neural-guided RANSAC: learning where to sample model hypotheses. In: IEEE/CVF International Conference on Computer Vision (ICCV), 27 October–2 November, Seoul, Korea (South), pp. 1–10 (2019)
89. Zhao, C., Cao, Z., Li, C., Li, X., Yang, J.: NM-net: mining reliable neighbors for robust feature correspondences. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 15–20 June 2019, Long Beach, CA, USA, pp. 1–10 (2019)
90. Zhang, J., et al.: Learning two-view correspondences and geometry using order-aware network. In: IEEE/CVF International Conference on Computer Vision (ICCV), 27 October–2 November 2019, Seoul, Korea (South), pp. 1–10 (2019)
91. Sun, W., Jiang, W., Trulls, E., Tagliasacchi, A., Yi, K.M.: ACNe: attentive context normalization for robust permutation-equivariant learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 14–19 June 2020, pp. 1–10 (2020)
92. Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperGlue: learning feature matching with graph neural networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 14–19 June 2020, pp. 1–10 (2020)
93. Rocco, I., Cimpoi, M., Arandjelović, R., Torii, A., Pajdla, T., Sivic, J.: Neighborhood consensus networks. In: Conference on Neural Information Processing Systems (NeurIPS), 2–8 December 2018, Montreal, Canada, pp. 1–12 (2018)
94. Tang, J., Kim, H., Guizilini, V., Pillai, S., Ambrus, R.: Neural outlier rejection for self-supervised keypoint learning. In: International Conference on Learning Representations (ICLR), 26 April–1 May 2020, pp. 1–14 (2020)
95. Rocco, I., Arandjelović, R., Sivic, J.: Efficient neighbourhood consensus networks via submanifold sparse convolutions. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12354, pp. 605–621. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-58545-7\\_35](https://doi.org/10.1007/978-3-030-58545-7_35)

96. Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: LoFTR: detector-free local feature matching with transformers. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 19–25 June 2021, pp. 1–10 (2021)
97. Zhou, Q., Sattler, T., Leal-Taixé, L.: Patch2Pix: epipolar-guided pixel-level correspondences. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 19–25 June 2021, pp. 1–10 (2021)
98. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **74**, 59–73 (2007)
99. Balntas, V., Lenc, K., Vedaldi, A.: A benchmark and evaluation of handcrafted and learned local descriptor. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017, Honolulu, HI, USA, pp. 1–10 (2017)
100. Song, W., Suganuma, M., Liu, X.: Matching in the dark: a dataset for matching image pairs of low-light scenes. ArXiv Preprint [arXiv:2109.03585](https://arxiv.org/abs/2109.03585) (2021)
101. Jin, Y., et al.: Image matching across wide baseline: from paper to practice. *Int. J. Comput. Vis.* **129**, 517–547 (2021)
102. Sattler, T., et al.: Benchmarking 6DOF outdoor visual localization in changing conditions. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 18–22 June 2018, Salt Lake City, UT, USA, pp. 1–10 (2018)
103. Toft, C., et al.: Long-term visual localization: revisited. *IEEE Trans. Pattern Anal. Mach. Intell.* (Early Access)
104. Zhou, Y., Qi, H., Ma, Y.: End-to-end wireframe parsing. In: IEEE/CVF International Conference on Computer Vision (ICCV), 27 October–2 November 2019, Seoul, Korea (South), pp. 1–10 (2019)
105. Dai, X., Yuan, X., Gong, H., Ma, Y.: Fully convolutional line parsing. ArXiv Preprint [arXiv:2104.11207](https://arxiv.org/abs/2104.11207) (2021)
106. Zhou, Y., Liu, S., Ma, Y.: NeRD: neural 3D reflection symmetry detector. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 19–25 June 2021, pp. 1–10 (2021)
107. Zhou, Y., et al.: HoliCity: a city-scale data platform for learning holistic 3D structures. ArXiv Preprint [arXiv: 2008.03286](https://arxiv.org/abs/2008.03286) (2020)
108. Salas-Moreno, R.F., Newcombe, R.A., Strasdat, H., Kelly, P.H.J., Davison, A.J.: SLAM++: simultaneous localisation and mapping at the level of objects. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 23–28 June 2013, Portland, OR, USA, pp. 1–8 (2013)
109. McCormac, J., Clark, R., Bloesch, M., Davison, A., Leutenegger, S.: Fusion++: volumetric object-level SLAM. In: International Conference on 3D Vision (3DV), 5–8 September 2018, Verona, Italy, pp. 1–10 (2018)
110. Sucar, E., Wada, K., Davison, A.: NodeSLAM: neural object descriptors for multi-view shape reconstruction. In: International Conference on 3D Vision (3DV), 25–28 November 2020, Fukuoka, Japan, pp. 1–10 (2020)