Evaluations on Matching Quality for 18 Different Descriptors and SIFT Keypoints over Various Inlier Ratios

Josef Maier[1], Martin Humenberger[1], Oliver Zendel[1], and Markus Vincze[2]

[1] AIT Austrian Institute of Technology
{josef.maier.fl, martin.humenberger,
oliver.zendel}@ait.ac.at
[2] Vienna University of Technology
vincze@acin.tuwien.ac.at

Abstract. In this document, we present additional results on the mean accuracy, precision, recall, and fall-out over various inlier ratios. The evaluations are performed on numerous datasets for different descriptors using SIFT [1] keypoints.

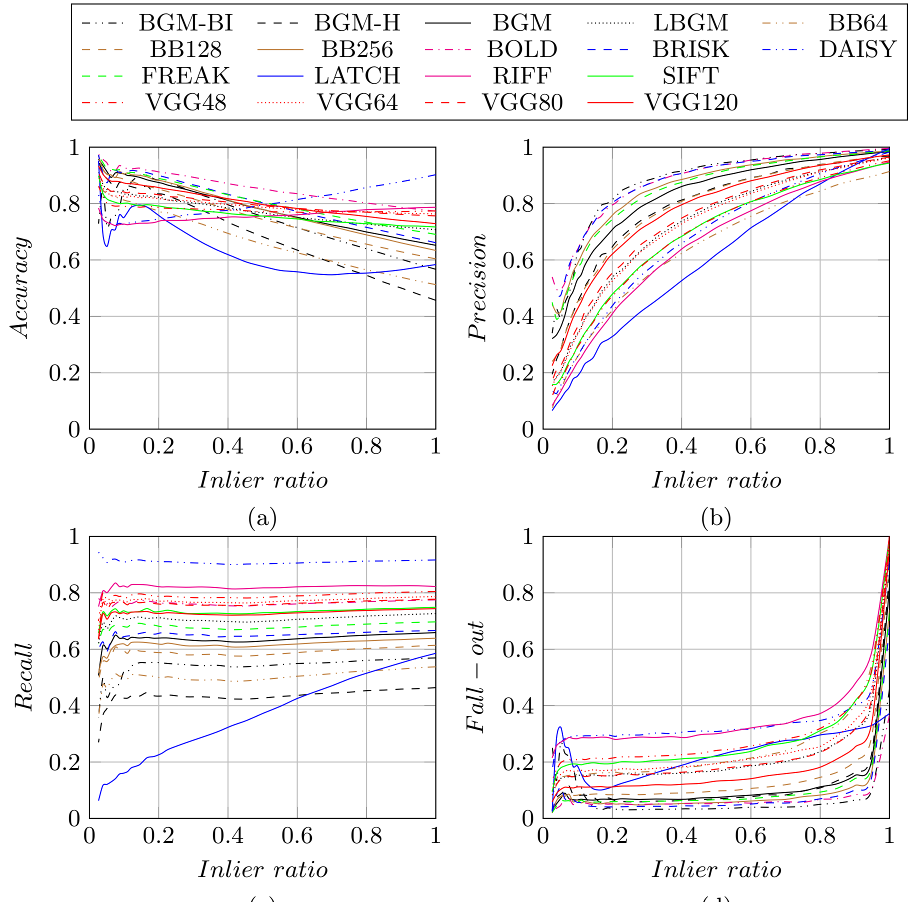

Fig. 1. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **KITTI disparity dataset** from Menze and Geiger [2]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

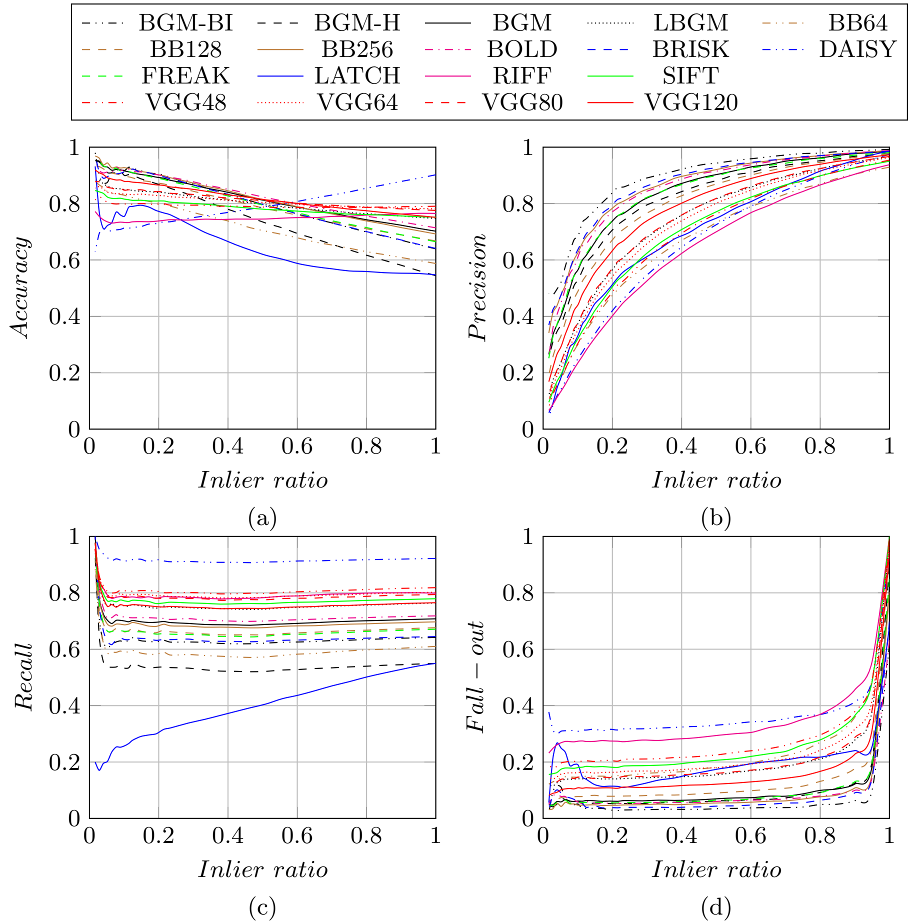

Fig. 2. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **KITTI flow dataset** from Menze and Geiger [2]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

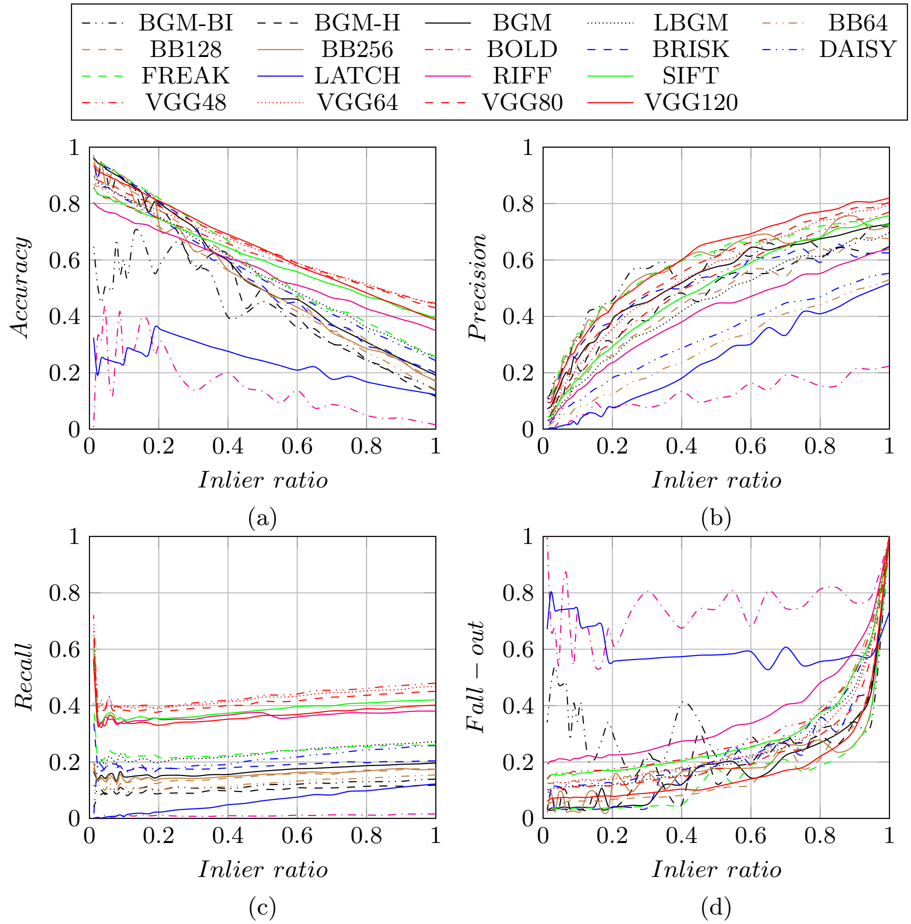

Fig. 3. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"graffiti" dataset** from Mikolajczyk *et al.* [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

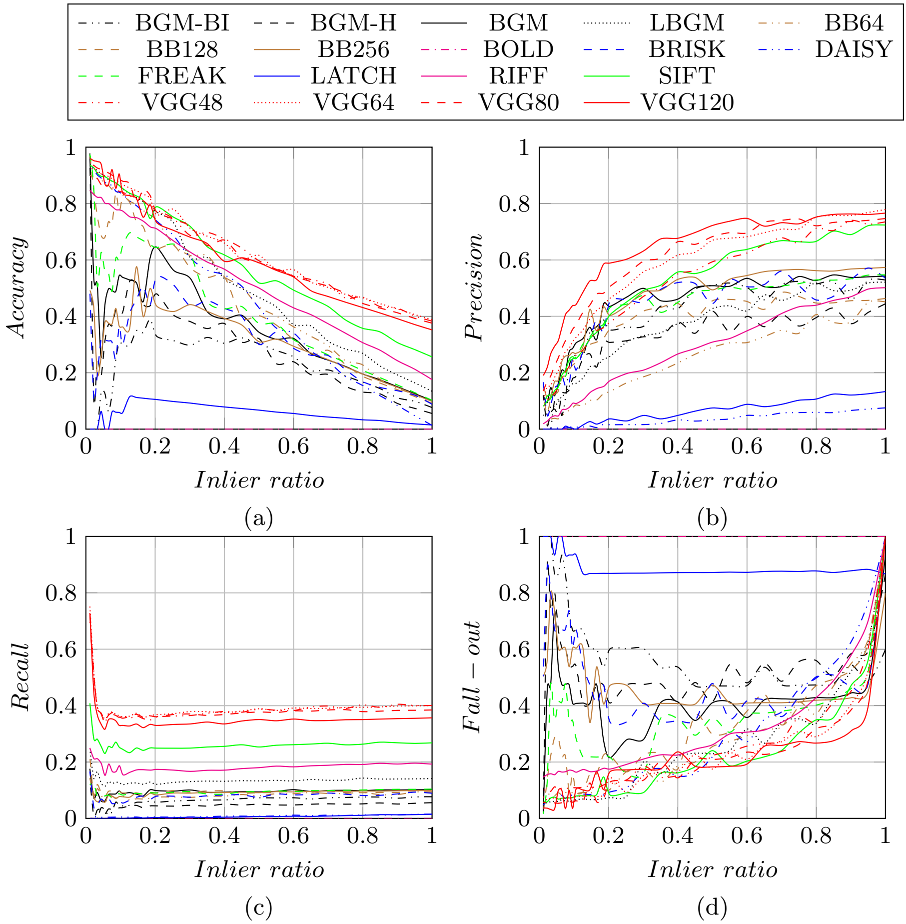

Fig. 4. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"bark" dataset** from Mikolajczyk *et al*. [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

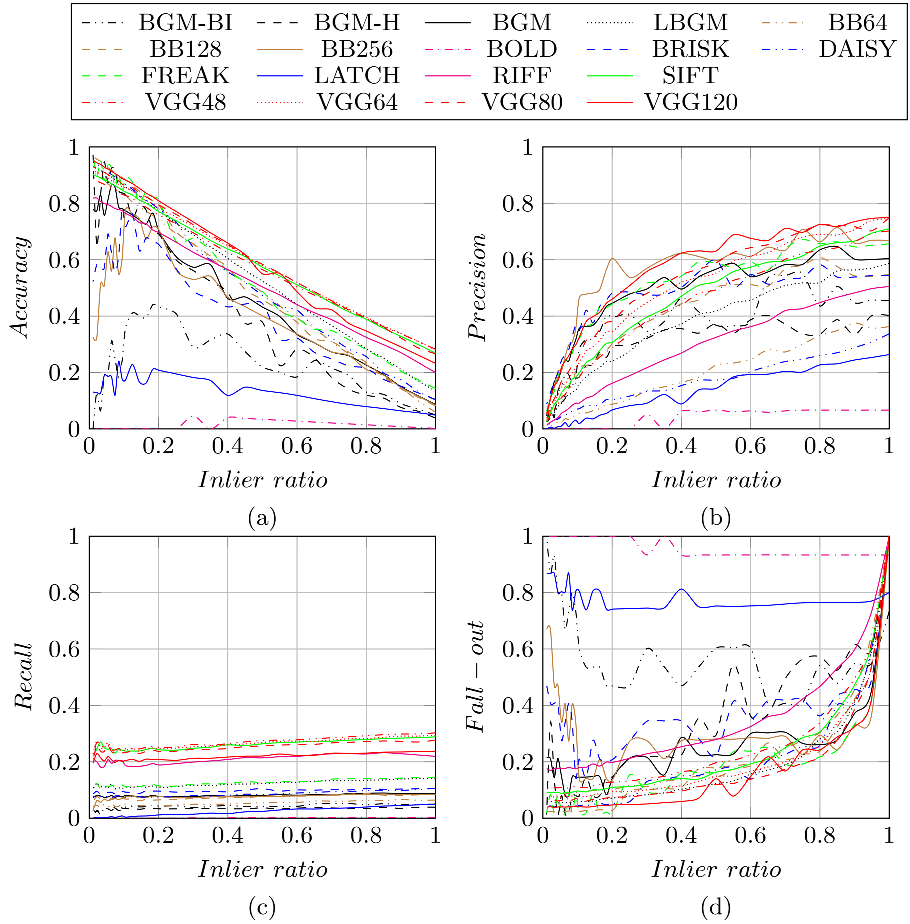

Fig. 5. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"boat" dataset** from Mikolajczyk *et al.* [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

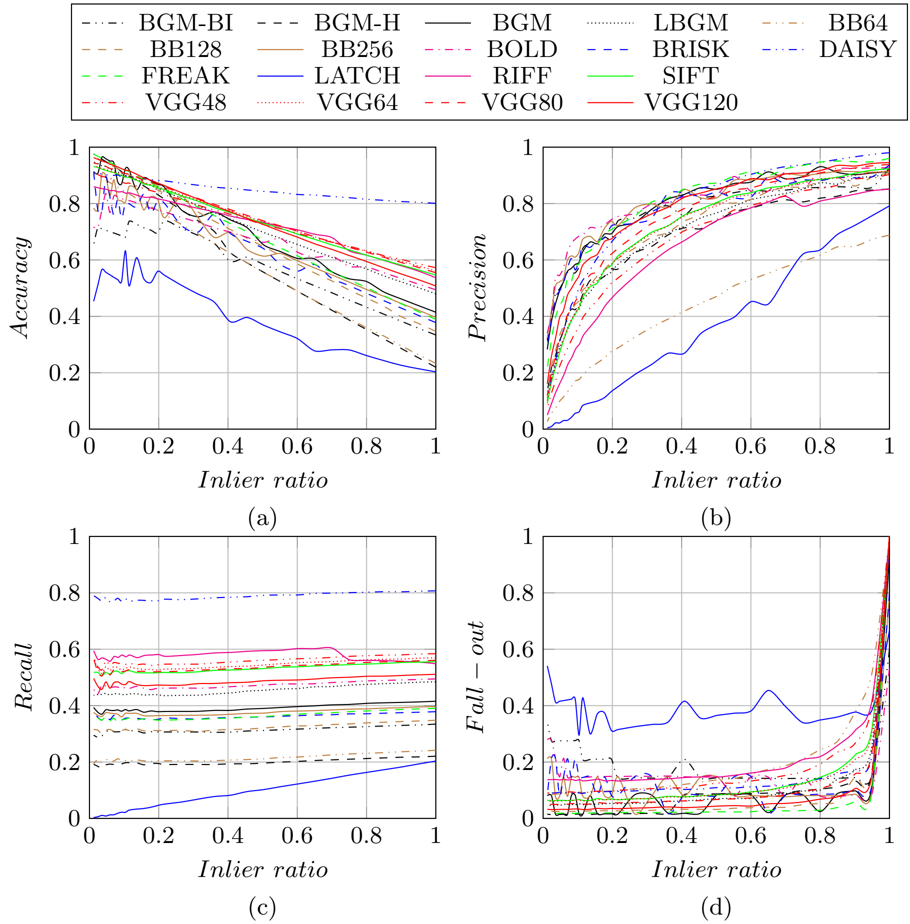

Fig. 6. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"wall" dataset** from Mikolajczyk *et al*. [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

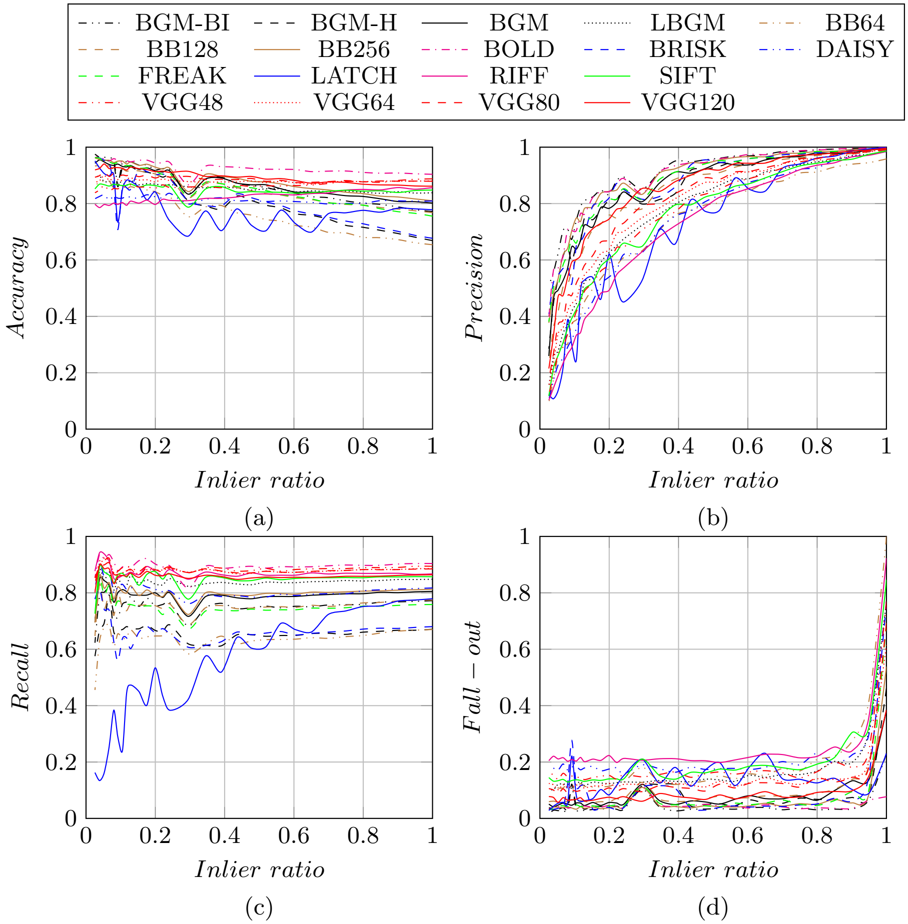

Fig. 7. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"bikes" dataset** from Mikolajczyk *et al.* [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

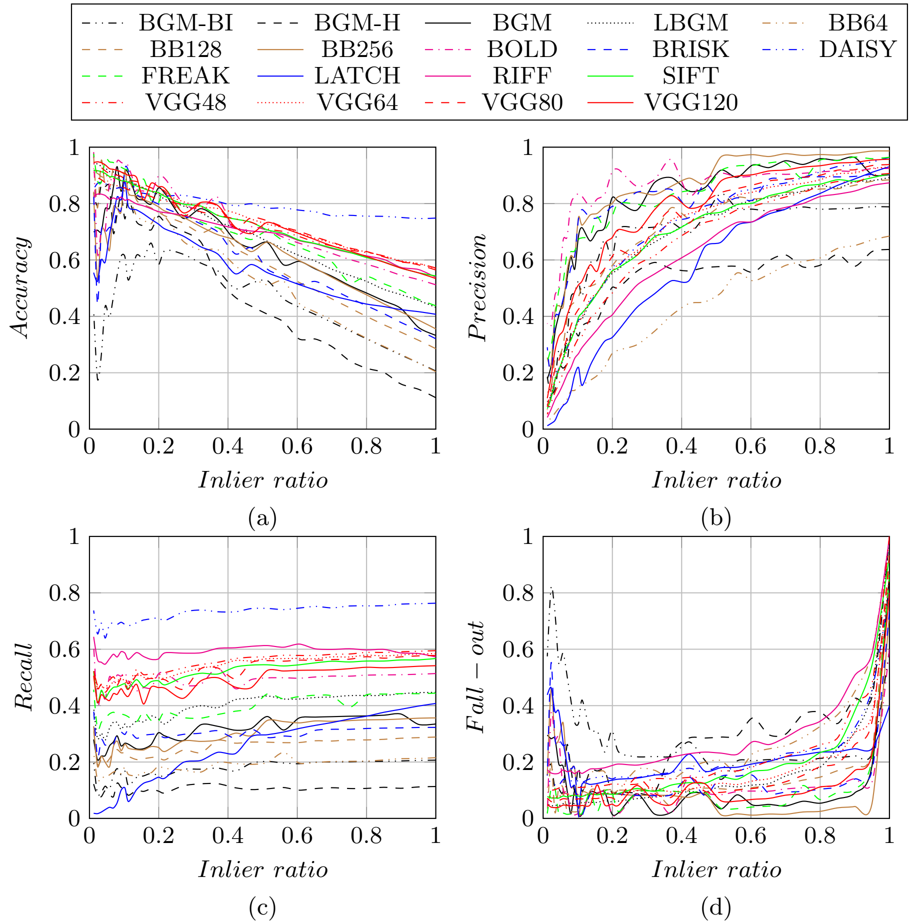

Fig. 8. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"JPEG" dataset** from Mikolajczyk *et al*. [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

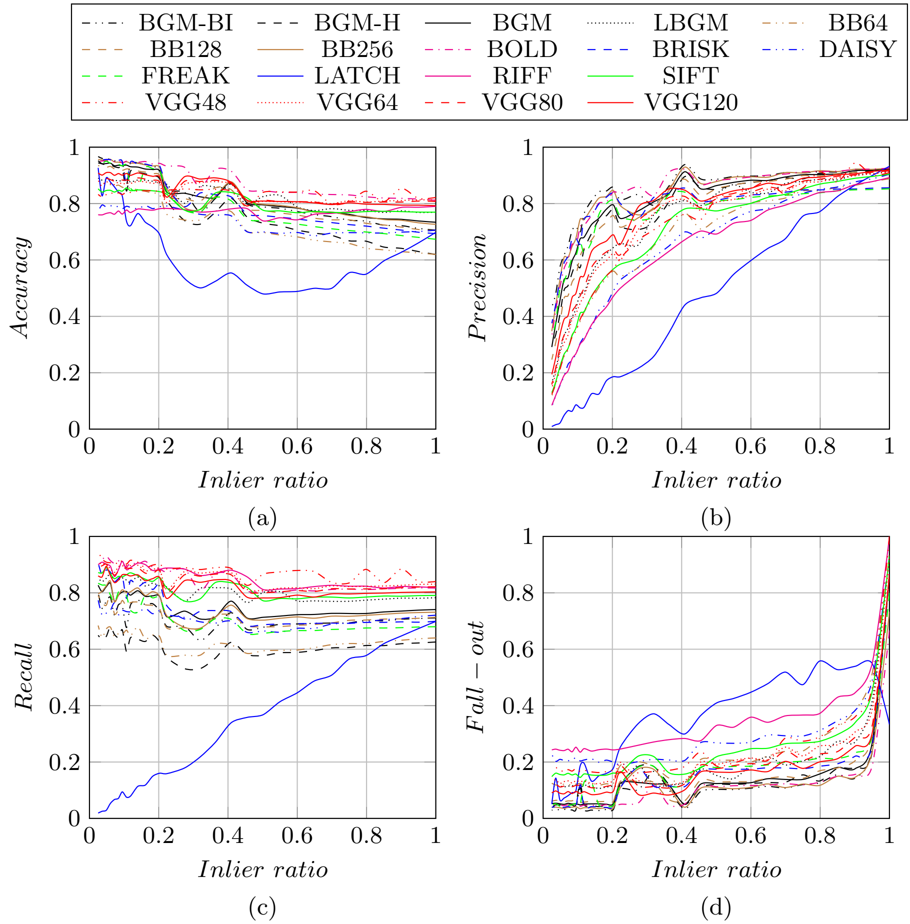

Fig. 9. Varying inlier ratio compared to mean (a) accuracy, (b) precision, (c) recall, and (d) fall-out using SIFT keypoints for the entire **"light" dataset** from Mikolajczyk *et al.* [13, 14]. For comparison, the following descriptors were used: BOLD [3], BRISK [4], DAISY [5], FREAK [6], LATCH [7], RIFF [8], SIFT [1], BGM-Bilinear (BGM-BI) [9], BGM-Hard (BGM-H) [9], BGM [9], LBGM [9], BinBoost [10, 11] with a descriptor size of 64 bits (BB64), 128 bits (BB128), and 256 bits (BB256), in addition to the VGG descriptor [12] with a descriptor size of 48 bits (VGG48), 64 bits (VGG64), 80 bits (VGG80), and 120 bits (VGG120). On the results of all algorithms, a ratio test was performed.

References

1. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**(2) (2004) 91–110
2. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 3061–3070
3. Balntas, V., Tang, L., Mikolajczyk, K.: Bold - binary online learned descriptor for efficient image matching. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2015) 2367–2375
4. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: Binary robust invariant scalable keypoints. In: International Conference on Computer Vision. (November 2011) 2548–2555
5. Tola, E., Lepetit, V., Fua, P.: Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(5) (May 2010) 815–830
6. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: Fast Retina Keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2012) 510–517
7. Levi, G., Hassner, T.: LATCH: learned arrangements of three patch codes. CoRR **abs/1501.03719** (2015)
8. Wu, S., Lew, M.S.: Riff: Retina-inspired invariant fast feature descriptor. In: Proceedings of the 22Nd ACM International Conference on Multimedia. MM '14, New York, NY, USA, ACM (2014) 1129–1132
9. Trzcinski, T., Christoudias, M., Lepetit, V., Fua, P.: Learning image descriptors with the boosting-trick. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. NIPS'12, USA, Curran Associates Inc. (2012) 269–277
10. Trzcinski, T., Christoudias, M., Fua, P., Lepetit, V.: Boosting binary keypoint descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, IEEE Computer Society (2013) 2874–2881
11. Trzcinski, T., Christoudias, M., Lepetit, V.: Learning image descriptors with boosting. IEEE Transactions on Pattern Analysis and Machine Intelligence **37**(3) (March 2015) 597–610
12. Simonyan, K., Vedaldi, A., Zisserman, A.: Learning local feature descriptors using convex optimisation. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(8) (August 2014) 1573–1585
13. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(10) (2005) 1615–1630
14. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A Comparison of Affine Region Detectors. International Journal of Computer Vision **65**(1-2) (2005) 43–72