Evaluations on Matching Quality for 133 Different Keypoint-Descriptor-Combinations over Various Inlier Ratios and Datasets

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 and datasets. The evaluations are performed for different keypoints and descriptors. For every keypoint-descriptor-combination the mean was calculated for every above stated quality parameter using all inlier ratios ranging from 1% to 100% and multiple datasets.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.714**	-	-	-	-
KAZE	-	**0.734**	-	-	-	-	-
ORB	-	0.557	0.570	0.569	0.576	0.684	**0.731**
LATCH	0.475	0.475	0.553	0.424	0.458	**0.568**	0.568
BOLD	0.576	**0.598**	0.594	0.541	0.598	0.573	0.565
BGM BILINEAR	**0.587**	0.556	0.565	0.470	0.510	0.527	0.547
BGM HARD	**0.652**	0.601	0.614	0.507	0.604	0.575	0.621
BINBOOST 64	**0.674**	0.671	0.665	0.640	0.639	0.649	0.662
BRISK	0.676	0.687	0.691	0.659	0.614	**0.751**	0.683
BINBOOST 256	**0.688**	0.637	0.649	0.583	0.589	0.645	0.649
RIFF	0.709	0.720	0.719	0.720	0.659	**0.761**	0.685
BGM	**0.716**	0.662	0.674	0.603	0.643	0.667	0.694
BINBOOST 128	**0.725**	0.683	0.687	0.620	0.673	0.673	0.698
DAISY	**0.736**	0.703	0.703	0.670	0.700	0.709	0.692
FREAK	0.741	0.732	0.730	0.740	0.687	**0.815**	0.716
LBGM	**0.746**	0.730	0.729	0.720	0.722	0.721	0.739
SIFT	**0.754**	0.745	0.741	-	0.709	0.714	0.534
VGG 48	0.761	0.759	0.757	**0.807**	0.755	0.796	0.773
VGG 64	0.778	0.772	0.765	**0.810**	0.760	0.801	0.781
VGG 80	0.784	0.774	0.767	**0.808**	0.756	0.807	0.784
VGG 120	0.786	0.771	0.764	0.784	0.717	**0.809**	0.776

Table 1. Mean accuracies for different descriptors and keypoint detectors. These mean accuracies are calculated over the mean accuracies for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1] in addition to the sequences "bark", "bikes", "boat", "graffiti", "JPEG", "light", and "wall" of the Oxford dataset [2]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ highest, ■ second, and ■ third highest mean descriptor accuracies for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.479**	-	-	-	-
KAZE	-	**0.627**	-	-	-	-	-
ORB	-	0.226	0.277	0.259	0.127	0.366	**0.439**
LATCH	0.199	0.232	**0.413**	0.219	0.135	0.371	0.382
BGM HARD	0.301	0.322	**0.327**	0.263	0.183	0.188	0.248
BINBOOST 64	0.333	**0.373**	0.367	0.355	0.231	0.284	0.286
BGM BILINEAR	0.367	0.392	**0.394**	0.327	0.253	0.276	0.314
BINBOOST 128	0.399	**0.437**	0.435	0.414	0.301	0.368	0.360
BRISK	0.414	0.463	0.460	0.456	0.264	**0.480**	0.457
BINBOOST 256	0.423	0.453	**0.457**	0.430	0.332	0.399	0.383
BGM	0.433	0.464	**0.469**	0.453	0.345	0.425	0.401
FREAK	0.446	0.507	0.499	0.520	0.305	**0.588**	0.543
BOLD	0.465	0.506	**0.508**	0.461	0.410	0.468	0.452
LBGM	0.488	0.521	0.525	**0.529**	0.407	0.476	0.466
VGG 120	0.562	0.609	0.612	0.655	0.480	**0.672**	0.571
SIFT	0.566	0.599	**0.619**	-	0.509	0.607	0.443
DAISY	0.569	**0.619**	0.619	0.599	0.499	0.569	0.542
RIFF	0.583	0.655	0.649	0.648	0.450	**0.678**	0.639
VGG 80	0.595	0.640	0.642	0.701	0.520	**0.706**	0.604
VGG 64	0.604	0.648	0.649	0.712	0.530	**0.713**	0.611
VGG 48	0.615	0.661	0.660	0.733	0.547	**0.737**	0.620

Table 2. Mean recall for different descriptors and keypoint detectors. These mean recall values are calculated over the mean recall values for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1] in addition to the sequences “bark”, “bikes”, “boat”, “graffiti”, “JPEG”, “light”, and “wall” of the Oxford dataset [2]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ highest, ■ second, and ■ third highest mean descriptor recall values for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.558**	-	-	-	-
KAZE	-	**0.526**	-	-	-	-	-
ORB	-	0.333	0.388	0.377	0.300	0.584	**0.626**
LATCH	0.344	0.329	0.526	0.322	0.304	**0.543**	0.475
BINBOOST 64	0.394	**0.412**	0.412	0.394	0.306	0.358	0.356
DAISY	0.438	0.415	0.419	**0.444**	0.393	0.411	0.383
RIFF	0.440	0.474	0.477	0.496	0.370	**0.521**	0.428
LBGM	0.520	0.518	0.521	**0.529**	0.469	0.518	0.496
VGG 48	0.523	0.537	0.538	**0.613**	0.507	0.583	0.529
SIFT	0.533	0.542	**0.552**	-	0.461	0.504	0.280
BOLD	**0.540**	0.496	0.503	0.500	0.519	0.518	0.483
BGM HARD	**0.541**	0.492	0.495	0.403	0.416	0.399	0.449
VGG 64	0.565	0.570	0.569	**0.640**	0.526	0.608	0.557
BINBOOST 128	**0.568**	0.543	0.558	0.530	0.484	0.555	0.525
VGG 80	0.588	0.590	0.589	**0.657**	0.542	0.629	0.572
VGG 120	0.632	0.621	0.622	**0.678**	0.567	0.669	0.602
BGM BILINEAR	**0.634**	0.559	0.576	0.487	0.540	0.551	0.542
BGM	**0.635**	0.585	0.609	0.571	0.559	0.617	0.598
BRISK	0.644	0.620	0.638	0.622	0.539	**0.745**	0.508
FREAK	0.645	0.631	0.636	0.648	0.552	**0.725**	0.508
BINBOOST 256	**0.672**	0.586	0.615	0.571	0.563	0.615	0.598

Table 3. Mean precision for different descriptors and keypoint detectors. These mean precision values are calculated over the mean precision values for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1] in addition to the sequences "bark", "bikes", "boat", "graffiti", "JPEG", "light", and "wall" of the Oxford dataset [2]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ highest, ■ second, and ■ third highest mean descriptor precision values for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.198**	-	-	-	-
KAZE	-	**0.261**	-	-	-	-	-
ORB	-	0.366	0.344	0.338	0.269	0.187	**0.150**
LATCH	0.419	0.455	0.367	0.495	0.414	**0.333**	0.352
BOLD	0.367	0.366	0.369	0.417	**0.313**	0.378	0.389
RIFF	0.287	0.304	0.302	0.301	0.295	**0.255**	0.349
BGM BILINEAR	**0.276**	0.344	0.327	0.437	0.346	0.323	0.333
DAISY	**0.236**	0.315	0.313	0.348	0.243	0.275	0.292
VGG 48	0.222	0.246	0.247	0.197	**0.181**	0.217	0.204
BINBOOST 64	0.210	0.232	0.234	0.282	**0.206**	0.223	0.210
SIFT	**0.205**	0.227	0.243	-	0.235	0.279	0.476
BRISK	0.195	0.209	0.199	0.243	0.220	**0.118**	0.238
BGM HARD	**0.191**	0.276	0.261	0.393	0.201	0.253	0.227
VGG 64	0.183	0.214	0.223	0.172	**0.160**	0.192	0.183
BINBOOST 256	**0.178**	0.275	0.253	0.332	0.281	0.231	0.225
LBGM	0.174	0.211	0.213	0.237	**0.153**	0.198	0.176
VGG 80	0.166	0.199	0.209	0.165	**0.158**	0.176	0.173
BGM	**0.158**	0.250	0.230	0.318	0.215	0.215	0.179
BINBOOST 128	**0.142**	0.224	0.209	0.297	0.164	0.193	0.168
VGG 120	**0.135**	0.178	0.190	0.166	0.190	0.146	0.159
FREAK	0.132	0.175	0.173	0.175	0.136	**0.091**	0.242

Table 4. Mean fall-out for different descriptors and keypoint detectors. These mean fall-out values are calculated over the mean fall-out values for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1] in addition to the sequences “bark”, “bikes”, “boat”, “graffiti”, “JPEG”, “light”, and “wall” of the Oxford dataset [2]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ lowest, ■ second, and ■ third lowest mean descriptor fall-out values for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.793**	-	-	-	-
KAZE	-	**0.748**	-	-	-	-	-
ORB	-	0.715	0.756	0.645	0.685	0.803	**0.811**
LATCH	0.677	0.708	**0.778**	0.552	0.681	0.769	0.771
BINBOOST 64	0.730	0.749	0.751	**0.755**	0.685	0.716	0.732
RIFF	0.748	0.763	0.760	0.761	0.705	**0.772**	0.682
BGM HARD	0.765	0.763	**0.772**	0.657	0.700	0.706	0.751
DAISY	0.776	0.755	0.755	**0.802**	0.718	0.763	0.728
SIFT	0.785	0.786	**0.797**	-	0.721	0.721	0.532
VGG 48	0.790	0.787	0.803	**0.835**	0.798	0.791	0.807
BGM BILINEAR	0.798	0.809	**0.818**	0.699	0.760	0.741	0.787
LBGM	0.806	0.807	0.813	**0.824**	0.785	0.779	0.810
BINBOOST 128	0.807	0.818	**0.821**	0.795	0.770	0.790	0.804
VGG 64	0.811	0.806	0.819	**0.847**	0.807	0.807	0.821
VGG 80	0.820	0.815	0.827	**0.855**	0.812	0.817	0.827
BRISK	0.831	**0.839**	0.832	0.739	0.768	0.807	0.754
VGG 120	0.832	0.831	0.841	**0.863**	0.817	0.837	0.838
BGM	0.833	0.837	**0.843**	0.821	0.800	0.816	0.828
BINBOOST 256	0.835	0.837	**0.842**	0.793	0.797	0.812	0.825
FREAK	0.838	0.844	0.840	0.808	0.776	**0.846**	0.748
BOLD	0.865	**0.866**	0.866	0.829	0.822	0.859	0.844

Table 5. Mean accuracies for different descriptors and keypoint detectors. These mean accuracies are calculated over the mean accuracies for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ highest, ■ second, and ■ third highest mean descriptor accuracies for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.632**	-	-	-	-
KAZE	-	**0.752**	-	-	-	-	-
ORB	-	0.424	0.516	0.412	0.247	0.565	**0.576**
LATCH	0.331	0.435	**0.595**	0.364	0.257	0.525	0.514
BGM HARD	0.491	0.526	**0.549**	0.469	0.298	0.336	0.411
BINBOOST 64	0.551	0.597	**0.604**	0.599	0.392	0.472	0.471
BGM BILINEAR	0.586	0.630	**0.649**	0.557	0.424	0.461	0.517
BINBOOST 128	0.631	0.679	**0.689**	0.671	0.502	0.577	0.572
BRISK	0.650	**0.694**	0.684	0.600	0.439	0.587	0.569
BINBOOST 256	0.657	0.696	**0.711**	0.675	0.541	0.605	0.603
BGM	0.672	0.711	**0.729**	0.716	0.563	0.639	0.621
FREAK	0.676	**0.726**	0.714	0.685	0.486	0.692	0.659
LBGM	0.734	0.763	0.783	**0.793**	0.644	0.694	0.698
BOLD	0.744	0.802	**0.804**	0.752	0.617	0.734	0.701
VGG 120	0.745	0.775	0.798	**0.815**	0.699	0.786	0.746
SIFT	0.756	0.772	**0.805**	-	0.703	0.746	0.598
VGG 80	0.776	0.797	0.823	**0.845**	0.738	0.815	0.771
VGG 64	0.785	0.803	0.829	**0.852**	0.748	0.822	0.777
VGG 48	0.800	0.816	0.840	**0.870**	0.764	0.839	0.785
RIFF	0.808	**0.842**	0.839	0.810	0.664	0.793	0.776
DAISY	0.916	0.941	0.941	**0.943**	0.827	0.912	0.864

Table 6. Mean recall for different descriptors and keypoint detectors. These mean recall values are calculated over the mean recall values for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ highest, ■ second, and ■ third highest mean descriptor recall values for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.681**	-	-	-	-
KAZE	-	**0.547**	-	-	-	-	-
ORB	-	0.532	0.632	0.540	0.496	**0.764**	0.713
RIFF	0.521	0.551	0.555	**0.563**	0.483	0.537	0.451
LATCH	0.531	0.585	**0.759**	0.521	0.509	0.727	0.641
BINBOOST 64	0.539	0.587	0.597	**0.603**	0.483	0.517	0.530
DAISY	0.543	0.538	0.543	**0.588**	0.498	0.524	0.490
VGG 48	0.569	0.582	0.602	**0.641**	0.589	0.555	0.574
SIFT	0.575	0.594	**0.606**	-	0.499	0.486	0.318
VGG 64	0.605	0.612	0.630	**0.669**	0.610	0.579	0.599
LBGM	0.617	0.629	0.639	**0.652**	0.615	0.570	0.608
VGG 80	0.625	0.631	0.647	**0.691**	0.626	0.596	0.613
VGG 120	0.669	0.674	0.686	**0.733**	0.660	0.642	0.648
BINBOOST 128	0.688	0.711	**0.717**	0.715	0.676	0.656	0.680
BGM HARD	0.702	0.705	**0.711**	0.624	0.615	0.588	0.651
BGM	0.739	0.742	**0.750**	0.745	0.733	0.691	0.723
FREAK	**0.753**	0.743	0.752	0.733	0.725	0.736	0.528
BINBOOST 256	**0.777**	0.774	0.777	0.754	0.757	0.710	0.752
BOLD	**0.786**	0.742	0.747	0.755	0.751	0.748	0.702
BRISK	**0.795**	0.772	0.779	0.728	0.780	0.787	0.562
BGM BILINEAR	**0.807**	0.799	0.797	0.709	0.789	0.719	0.770

Table 7. Mean precision for different descriptors and keypoint detectors. These mean precision values are calculated over the mean precision values for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ highest, ■ second, and ■ third highest mean descriptor precision values for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

Descr. / Keypoint	SIFT	KAZE	AKAZE	MSD	FAST	BRISK	ORB
AKAZE	-	-	**0.153**	-	-	-	-
KAZE	-	**0.305**	-	-	-	-	-
ORB	-	0.223	0.175	0.274	0.163	**0.091**	0.098
DAISY	0.342	0.389	0.392	**0.312**	0.386	0.358	0.390
RIFF	0.332	0.329	0.331	0.317	0.323	**0.292**	0.414
VGG 48	0.265	0.278	0.266	**0.223**	0.230	0.282	0.232
SIFT	**0.248**	0.254	0.257	-	0.323	0.347	0.549
BINBOOST 64	0.223	0.211	0.209	0.204	0.209	0.211	**0.187**
VGG 64	0.221	0.239	0.230	**0.188**	0.205	0.247	0.204
LATCH	0.202	0.194	0.118	0.345	0.160	**0.111**	0.121
VGG 80	0.201	0.217	0.211	**0.166**	0.189	0.225	0.189
LBGM	0.198	0.212	0.212	0.198	**0.177**	0.226	0.179
VGG 120	0.159	0.174	0.170	**0.131**	0.156	0.175	0.156
BINBOOST 128	0.127	0.131	0.129	0.149	0.113	0.133	**0.111**
BGM HARD	0.103	0.119	0.119	0.240	0.102	0.126	**0.100**
BGM	0.102	0.112	0.114	0.127	0.094	0.117	**0.092**
FREAK	0.094	0.116	0.115	0.140	**0.093**	0.102	0.252
BINBOOST 256	0.085	0.095	0.097	0.139	0.083	0.105	**0.081**
BGM BILINEAR	0.081	0.084	0.084	0.200	**0.064**	0.113	0.076
BRISK	0.079	0.097	0.101	0.179	**0.071**	0.092	0.192
BOLD	**0.076**	0.110	0.114	0.132	0.089	0.093	0.107

Table 8. Mean fall-out for different descriptors and keypoint detectors. These mean fall-out values are calculated over the mean fall-out values for all inlier ratios ranging from 1% to 100% and the datasets KITTI 2015 flow and disparity [1]. The bold values mark the best performing detector-descriptor combination for every descriptor. The colored cells mark the ■ lowest, ■ second, and ■ third lowest mean descriptor fall-out values for every detector. Some cells are empty due to a incompatible keypoint-descriptor combination.

References

1. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 3061–3070
2. 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