Which is which? Evaluation of local descriptors for image matching in real-world scenarios

Which is Which? Evaluation of local descriptors for image matching in real-world scenarios

Fabio Bellavia[0000−0002−1688−8476] and Carlo Colombo[0000−0001−9234−537X]

CVG/DINFO, University of Florence, via di S. Marta 3, 50134, Firenze, Italy
{fabio.bellavia,carlo.colombo}@unifi.it
http://cvg.dsi.unifi.it/wisw.caip2019

Abstract. Matching with local image descriptors is a fundamental task in many computer vision applications. This paper describes the WISW contest held within the framework of the CAIP 2019 conference, aimed at benchmarking recent descriptors in challenging planar and non-planar real image matching scenarios. According to the contest results, the descriptors submitted to the competition, most of which based on deep learning, perform significantly better than the current state-of-the-art in image matching. Nonetheless, there is still room for improvement, especially in the case of non-planar scenes.

Keywords: Local image descriptors · Image matching · Deep descriptors.

1 Introduction

Local image descriptors [13] play a critical role in establishing reliable point correspondences among several images in many computer vision applications, such as image stitching [8], 3D reconstruction [32] and visual odometry [15]. Research on this topic is still very active today. Impressive advances have been obtained in the last few years both with handcrafted and data-driven descriptors, thanks to careful modeling and design strategies, deep learning architectures, big data and efficient hardware.

The "Which is Which? Evaluation of local descriptors for image matching in real-world scenarios" (WISW) contest, held within the framework of the CAIP 2019 conference, was aimed at benchmarking recent descriptors in challenging real image matching scenarios, facing with both planar and non-planar scenes. This paper reports the rationale, setup protocols and datasets employed in the contest, and comparatively analyzes the results achieved by the competing descriptors, also in relation to the state-of-the-art in the field. There were seven different submissions by four distinct research groups. The submitted descriptors were all brand new and in some cases still unpublished. According to the results, the competing local image descriptors, although designed as variants of previous approaches, generally showed remarkable improvements with respect to the state-of-the-art. Descriptors based on deep learning showed to achieve the most important enhancements.

The paper is organized as follows. Motivation and related work on local image descriptor benchmarks are presented in Sec. 2. Datasets and setup protocols for both the planar and non-planar scenarios are defined in Sec. 3. Baseline and submitted descriptors are described in Sec. 4. Results are discussed in Sec. 5, and conclusions are outlined in Sec. 6, together with some directions for future work.

2 Motivation and related work

The factors that affect the matching accuracy and robustness of local image descriptors include the scene content, the image transformations involved, and the requirements in terms of computational efficiency (both in space and time). Adaptability to non-planar scene content and relevant viewpoint changes are the most important properties that a good descriptor must have in order to be used in general, real-world image matching applications.

The most consolidated benchmarks on local image descriptors contemplate planar scenes only and are based on the standard Oxford evaluation protocol [21, 23]. The recent HPatches [2] is perhaps the most representative planar benchmark. In HPatches, ground-truth matches are estimated according to the overlap error, that can be obtained without ambiguity using image patches and their homography-based reprojections. Local patches are preferred over images as input, as they limit the influence of factors other than the descriptor itself on matching performance. Following such protocol, the planar evaluation case of the WISW benchmark also uses patches as input. On the other hand, differently from HPatches, custom patch orientations are allowed, since these are an integral part of the descriptor. Moreover, besides viewpoint transformations, typically considered by HPatches, their combinations with other illumination changes, blur and noise effects are also considered in WISW (illumination changes are also benchmarked by HPatches, but separately from viewpoint changes). At any rate, evaluation on planar scenes provides only a limited insight into descriptors since, for instance, it is not able to analyze and investigate the accuracy in the presence of self-occlusions in a real 3D scene.

In order to overcome the limitations of planar scenes, non-planar scenes have also been used in recent benchmarks. For this purpose, ground-truth was directly estimated (a) using stereo matching [16] or Structure-from-Motion [30], (b) through complex sensor-based system setups [12, 33], or (c) according to some approximation scheme [6, 26]. Alternative benchmarks were also proposed, that characterize indirectly matching robustness (d) by checking the correctness of the output for a given specific application task, such as object retrieval [14] or visual odometry [7]. However, none of these approaches is without drawbacks, since ground-truth may not be available for some image region (a,b), it can be erroneously estimated (a,c,d), or it can be biased towards the considered application (a,d). In WISW, the benchmark for non-planar scenes first introduced in [3], and based on a piecewise approximation of the overlap error, is used. As shown in Sec. 3.2, this benchmark is a natural extension of the planar bench-

mark, it can always provide ground-truth data with a low false positive rate, and it is not biased towards any specific application.

3 Benchmark setup

Figure 1 shows the patch extraction pipeline adopted for WISW. There are several aspects that can influence the evaluation, including the keypoint detector employed, the patch normalization strategy, the distance used to compare the descriptors, and the matching strategy by which to assign putative correspondences. In order to have a fair comparison without ambiguities, for the contest most the above factors were fixed in advance. The HarrisZ [5] detector was used to extract affine patches (Fig.1a), that were then normalized into circles with radius of $48 \times \sqrt{2} \approx 68$ pixels, extending by a factor of $\sqrt{2}$ the original normalized circular region of 48 pixels radius (Fig.1b). Pixels outside the extended circular patch were masked. Contest participants could optionally assign their own orientation to the patch, by computing it on the 97×97 square patch (marked in white in Fig.1b) inside the extended patch ($97 = 1 + 2 \times 48$). A default orientation for each patch was provided, computed by the deep learning approach described in [40] trained on EdgeFoci [41] patches. The extended and rotated circular patches (Fig.1c) were then cropped into the final 97×97 square patches (Fig.1d), and provided as input to descriptors.

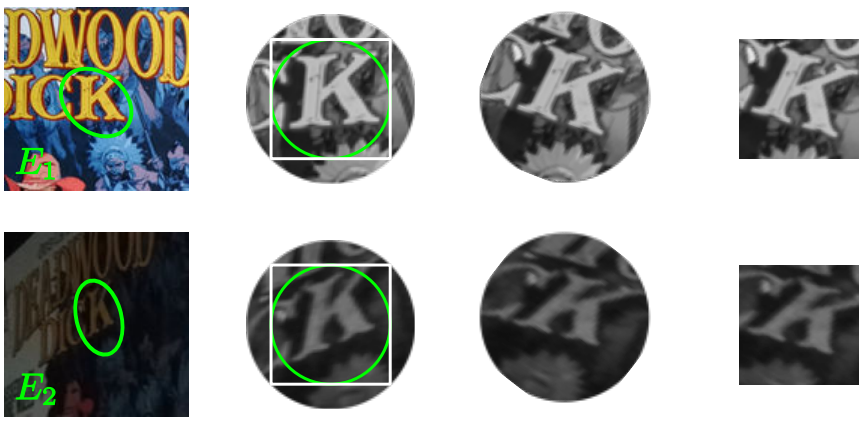


Fig. 1. Patch extraction process for two corresponding keypoints. Please refer to the text for details (best viewed in color).

For a given pair of images, a matrix representing the distance between all the keypoint pairs of the two images was generated by the contest participants.

Such matrix reflects the kind of distance employed by the descriptor, but can also allow one to exploit descriptor distance statistics inside images, as done in [4]. Finally, the distance table was employed to extract the best matches according to their distance in a greedy way, so as to avoid that two matches share a common keypoint. The matches were then ordered according to the Nearest Neighbor Ratio (NNR) [19]. Since NNR is asymmetric and depends on which image is taken as reference, a symmetric version considering the average between the two possible choices was used. Using the proposed workflow, descriptor input is fixed as for HPatches [2] but, differently from it, the definition of custom orientations is allowed.

Notice that the WISW benchmark does not consider running times and computational efficiency, since these parameters are strongly dependent on the hardware and software implementations (e.g. CPU, SIMD, GPU). Moreover, emerging deep learning approaches to image matching that bind together keypoint detector and descriptor [9, 28, 39] are excluded from the comparison, since the proposed benchmark fixes the keypoint detector to focus only on local image descriptor behavior.

3.1 Planar scenes

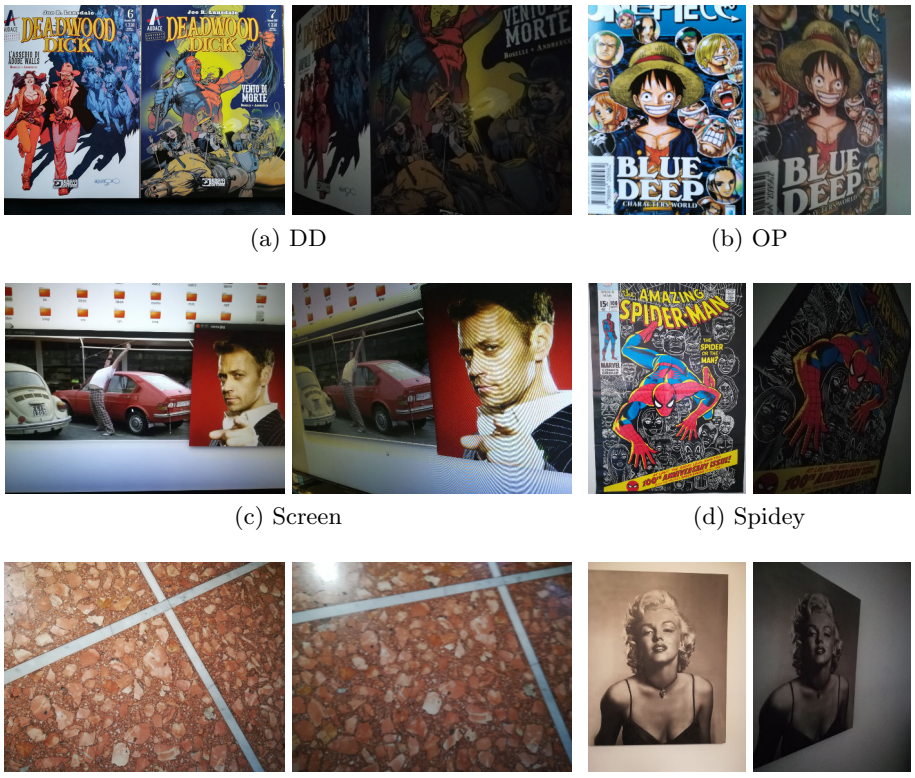
Dataset. The dataset employed consists of 15 different scenes of 6 images each, for a total of $15 \times (6-1) = 75$ image pairs. The scenes include "Bark", "Boat", Graffiti" and "Wall" from the Oxford dataset [22], the whole Viewpoint dataset [40] and 6 new scenes (see Fig. 2). In addition to viewpoint changes, the new scenes also include at the same time illumination changes, blur and Moiré pattern noise, thus increasing the complexity of the image transformations at hand.

Evaluation protocol. Planar scene evaluation follows the protocol described in [21]. The overlap error, computed according to [18] (i.e., without employing discrete approximations as in [21]), is used to define ground-truth matches. A match is considered correct if the overlap error between the elliptical keypoint region on the reference image and the reprojection of the elliptical keypoint region on the other image through the homography relating the viewpoint transformation is less than 50% (see Fig. 3). Finally, for each input image pair, the mean Average Precision (mAP) is computed from the precision/recall curve, interpolating data as described in [10].

3.2 Non-planar scenes

Dataset. The dataset for the non-planar case contains images from 35 different scenes used in other works (19 having 3 images, the remaining 16 with 2 images only), for a total of $19 \times (3-1) + 16 \times (2-1) = 73$ image pairs. The dataset extends the one used in [3], which included only 42 image pairs (see the first two rows of Fig. 4 for some examples). Notice that in addition to the images to probe,

this dataset contains reference matches that are employed both for computing the approximated overlap error and for refining correspondences according to their local flow [3].



(a) DD (b) OP

(c) Screen (d) Spidey

(e) Floor (f) Marilyn

Fig. 2. Sample image pairs from the new six planar scenes included in the benchmark (best viewed in color).

Evaluation protocol. Non-planar scene evaluation follows the protocol described in [3], employing the approximated overlap error defined in [6] for computing the ground-truth. Unlike other non-planar benchmarks, the approximated overlap relies on the whole local descriptor patch and not on the keypoint position only, thus being a natural extension of the overlap error to the non-planar case. This benchmark was shown to give a very low false positive rate (less than 5%), which does not affect descriptor ranking in unsupervised evaluations [6]. Similarly to the planar case, the approximated overlap error threshold was set

to 50%. Unlike the planar case, the number of total correct matches employed to compute the recall denominator is not established by considering all the possible keypoint pair combinations, but only the union set of all the putative matches output by all the descriptors included in the contest. This is done to reduce false positive matches in the estimation of the total number of matches. It was verified that this way to compute the mAP does not change the relative rank between descriptors, when applied to the planar case, although mAP values may slightly differ when some descriptors are removed from or added to the evaluation (compare columns ▱ and ▨ in Table 1 later in the evaluation). Example of correct matches according to the approximated overlap error are shown in the last row of Fig. 4.

(a) (b)

Fig. 3. Planar scene evaluation. (a) Overlap error computation for the two keypoints E_1 and E_2 in Fig. 1, belonging to the image pair in Fig. 2a. The elliptical keypoint region E_2 is reprojected into E'_2 through the viewpoint homography and the overlap error is computed as $1-(E_1 \cap E'_2)/(E_1 \cup E'_2)$. (b) Flow lines for correct matches among all those evaluated in the contest, different color intensities correspond to different overlap error values (best viewed in color and zoomed in).

4 Local image descriptors under evaluation

Seven local image descriptors were submitted to WISW. These include **SOSNet** [35], still unpublished at contest time, the recent **HardNet**$_A$ [29], obtained by training HardNet [24] on AMOS [29] and other datasets, **RalNet Shuffle** using the RalNet architecture [38] and additionally cropping and shuffling patches at training time, and **RsGLOH2**, "square rooting" sGLOH2 [4] according to RootSIFT [1]. Two variants of HardNet$_A$, exploiting the deep networks described

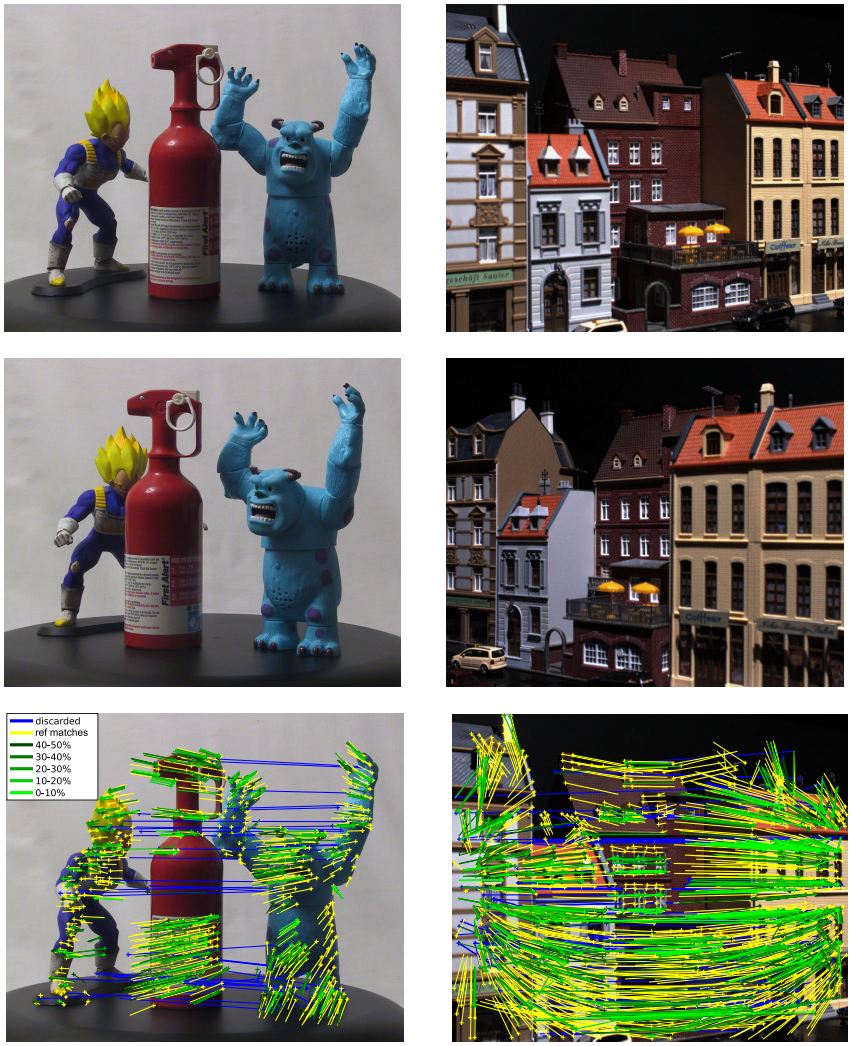


Fig. 4. (1^{st}, 2^{nd} rows) First and second images for two non-planar scene pairs included in the benchmark. (3^{th} row) Flow lines for correct matches among all those evaluated in the contest, together with flow of reference matches and of matches discarded by local flow heuristic. Different color intensities correspond to different approximated overlap error values (best viewed in color and zoomed in).

in [25] either for custom orientation assignment or to accommodate patches before extracting the descriptor, were also submitted as **OriNet+HardNet**$_A$ and **AffNet+HardNet**$_A$, respectively. The contest also included a variant of BisGLOH2 [4], named **BisGLOH2***, using more rotations at matching time

than the default ones. With the exception of the handcrafted RsGLOH2 and BisGLOH2*, all the submitted descriptors were data-driven deep descriptors.

In addition to the descriptors submitted to the contest, several recent state-of-the-art descriptors were included as baseline, for a total of 22 descriptors. These include seven deep descriptors, i.e. **GeoDesc** [20], **DOAP** [17] and **L2Net** [34] together with their binary versions, **HardNet** [24], and **DeepDesc** [31], three other kinds of data-driven descriptors, i.e. **MIOP** [36] and **RFD** [11] (in both its two variants), and five handcrafted descriptors, i.e., **RootSIFT** [1], **MKD** [27], **LIOP** [37], **sGLOH2** [4] (in both its regular and binary versions). For baseline descriptors, their publicly available implementations were employed.

5 Evaluation results

Table 1 reports the mAP in the case of planar and non-planar scenes ( and  columns, respectively), averaged on all the image pairs of the datasets, together with the main descriptor properties. Detailed mAPs for each image pair can be found online[1], together with the code and data used in the WISW benchmark, freely available for reproducibility and further comparisons on future local image descriptors.

SOSNet is the best performing descriptor on both the planar and non-planar cases. The results of HardNet$_A$ and its variants, that follow in the ranking, also offer clear insights about the impact of training data and patch normalization in the matching process. HardNet$_A$ significantly improves on HardNet by simply employing a better training set. At the same time, affine patch accommodation thanks to AffNet preprocessing appears to be very suitable for non-planar scenes, although it slightly worsens the results in the planar case. This is quite reasonable, since being able to tolerate more patch transformations unavoidably decreases the discrimination power. Concerning OriNet [25], the default patch orientation system of the deep network detailed in [40] seems to be slightly better, possibly due to the difference between the keypoint detector employed during training and that used to generate input patches. RsGLOH2 achieves the best results among the handcrafted descriptors, while RalNet Shuffle and BisGLOH2* are comparable with average baseline descriptors.

Considering baseline descriptors only, the recent GeoDesc achieves the best results, followed by HardNet and L2Net. With our benchmark, HardNet obtains slightly better results than L2Net on planar scenes and slightly worse on non-planar scenes. sGLOH2, BisGLOH2, binary L2Net and DOAP follow next. As for L2Net and HardNet, sGLOH2 and BisGLOH2 are better than DOAP on non-planar scenes and worse on planar scenes. MKD and RootSIFT come next, followed by the remaining descriptors.

With respect to a recent evaluation using a very similar setup protocol [3], some differences in the descriptor relative rank can be noted (e.g. RootSIFT in WISW evaluation is better than RFD, as opposed to what reported in [3]).

[1] `https://drive.google.com/open?id=1P1easA8UwmFyAVYzu2K4tk4zu88Jg4Px`

Table 1. Contest evaluation results

		mAP (%)						info					
		▱	▨	▥	▣	▢	▤	✓	✓	✓		#	type
L_2	SOSNet	76.30	77.58	74.01	53.40	54.73	60.76	✓			[35]	128	`float`
	AffNet+HardNet$_A$	74.11	75.09	71.71	52.34	53.64	59.98	✓			[25, 29]	128	`uchar`
	OriNet+HardNet$_A$	73.50	74.38	71.14	49.92	51.22	57.09	✓			[25, 29]	128	`uchar`
	HardNet$_A$	74.29	75.22	72.14	50.08	51.38	57.47	✓			[29]	128	`uchar`
	GeoDesc	75.60	76.67	71.83	47.56	48.78	55.47				[20]	128	`uchar`
	HardNet	71.49	72.17	68.86	47.80	49.01	55.37				[25, 29]	128	`uchar`
	L2Net	69.49	70.20	66.97	48.79	50.05	56.46				[34]	256	`float`
	RalNet Shuffle	65.51	66.50	62.76	41.53	42.62	49.75	✓			[38]	128	`uchar`
	DOAP	69.80	70.57	67.19	40.66	41.77	44.99				[17]	128	`float`
	MIOP	56.83	57.49	52.13	33.38	34.24	39.33			✓	[36]	128	`float`
	DeepDesc	53.24	53.90	56.32	37.03	38.02	44.93				[31]	128	`float`
L_1	RsGLOH2	70.68	72.50	67.84	48.19	49.48	56.11	✓	✓	✓	[4]	256	`float`
	sGLOH2	67.25	69.59	63.50	44.86	46.08	52.49		✓	✓	[4]	256	`uchar`
	RootSIFT	58.46	59.25	56.74	37.73	38.73	44.77		✓		[1]	128	`uchar`
	LIOP	54.51	54.97	49.50	32.05	32.91	37.93		✓	✓	[37]	144	`uchar`
H	BisGLOH2*	66.80	68.01	62.33	44.18	45.40	51.76	✓	✓	✓	[4]	1152	`bit`
	BisGLOH2	66.04	66.99	62.27	44.08	45.29	51.63		✓	✓	[4]	1152	`bit`
	Binary L2Net	63.11	63.96	61.06	43.33	44.47	50.86				[34]	256	`bit`
	Binary DOAP	54.24	54.82	52.74	34.57	35.49	41.41				[17]	128	`bit`
	RFD$_G$	53.58	53.99	50.75	34.17	35.06	40.40				[11]	406	`bit`
	RFD$_R$	52.62	53.22	50.28	32.96	33.85	39.31				[11]	293	`bit`
$*$	MKD	59.52	60.42	56.40	39.05	40.09	45.70				[27]	128	`float`

▱ planar ▨ with alternative recall computation ▥ "viewpoint only" dataset of [3]
▣ non-planar ▢ removing image pairs with mAP<5% ▤ non-planar dataset of [3]
✓ contest submission ✓ hand-crafted ✓ rotationally invariant ✓ references # vector length
L_2 Euclidean distance L_1 Manhattan distance H Hamming distance $*$ dot product

More than to the different number of image pairs evaluated (almost doubled in WISW), this is due to a better input patch registration and to the final matching strategy in the proposed benchmark (mAP results on the same planar and non-planar scene datasets of [3] are reported as columns ▥ and ▤ in the Table 1, respectively). This underlines a critical issue when designing the matching pipeline, since descriptors more tolerant to inaccurate patch registration can be less discriminative. Finally, in order to consider possible inaccuracies in the ground-truth estimation due to false positives in the non-planar scenes, average mAP excluding scenes with very low mAP have been computed, but no relevant changes in descriptor rank were observed (compare columns ▣ and ▢ in the Table 1, where only two image pairs were removed).

6 Conclusions and future work

This paper presented the results of the WISW contest, held within the framework of the CAIP 2019 conference, aimed at benchmarking recent local descriptors in challenging real image matching scenarios. For this purpose, descriptors were

evaluated on both planar and non-planar scenes, since relevant viewpoint changes and adaptability to non-planar objects and self-occlusions in the scene represent the most general and significant real-world environments. The WISW contest extended existing datasets by adding more test images. In the case of planar scenes, viewpoints changes were combined with other image transformations such as illumination variations and blur, so as to achieve a more realistic and challenging complexity.

Evaluation results showed remarkable improvements of recent descriptors with respect to the state-of-the-art. These were particularly impressive for some of the descriptors based on deep learning, thanks to their smart architecture, combined to the ever increasing availability of big data and modern hardware capabilities. Nevertheless, there is still room for improvement, especially in the case of non-planar scenes.

The proposed benchmark evidenced the fact that, beside descriptors, other factors often overlooked, such as patch normalization and matching strategy, are critical for image matching and are worth to be better investigated in the future. Future work will also include the evaluation of novel descriptors in the benchmark. To this aim, **the WISW contest will remain permanently open**. Furthermore, image pairs with increased complexity combining more image transformations simultaneously will be added in the datasets and, in the case of non-planar scenes, refined extensions of the approximated overlap error will be investigated to further reduce the number of false positive matches.

Acknowledgment

The Titan Xp used for this research was generously donated by the NVIDIA Corporation.

References

1. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2911–2918 (2012)
2. Balntas, V., Lenc, K., Vedaldi, A., Mikolajczyk, K.: HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3852–3861 (2017)
3. Bellavia, F., , Colombo, C.: An evaluation of recent local image descriptors for real-world applications of image matching. In: Proceedings of the IAPR International Conference on Machine Vision Applications (MVA) (2019)
4. Bellavia, F., Colombo, C.: Rethinking the sGLOH descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(4), 931–944 (2018)
5. Bellavia, F., Tegolo, D., Valenti, C.: Improving Harris corner selection strategy. IET Computer Vision **5**(2), 86–96 (2011)
6. Bellavia, F., Valenti, C., Lupascu, C.A., Tegolo, D.: Approximated overlap error for the evaluation of feature descriptors on 3D scenes. In: Proceedings of the International Conference on Image Analysis and Processing (ICIAP). pp. 270–279 (2013)

7. Bian, J., Zhang, L., Liu, Y., Lin, W.Y., Cheng, M.M., Reid, I.D.: MatchBench: An evaluation of feature matchers. In: arXiv (2018)
8. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision **74**(1), 59–73 (Aug 2007)
9. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2018)
10. Everingham, M., Van Gool, L., Williams, C., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. vol. 88, pp. 303–338 (2010)
11. Fan, B., Kong, Q., Trzcinski, T., Wang, Z., Pan, C., Fua, P.: Receptive fields selection for binary feature description. IEEE Transactions on Image Processing **26**(6), 2583–2595 (2014)
12. Fan, B., Kong, Q., Wang, X., Wang, Z., Xiang, S., Pan, C., Fua, P.: A performance evaluation of local features for image-based 3d reconstruction. IEEE Transactions on Image Processing (2019)
13. Fan, B., Wang, Z., Wang, F.: Local Image Descriptor: Modern Approaches
14. Fan, B., Wu, F., Hu, Z.: Rotationally invariant descriptors using intensity order pooling. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(10), 2031–2045 (2012)
15. Fanfani, M., Bellavia, F., Colombo, C.: Accurate keyframe selection and keypoint tracking for robust visual odometry. Machine Vision and Applications **27**(6), 833–844 (2016)
16. Fraundorfer, F., Bischof, H.: A novel performance evaluation method of local detectors on non-planar scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). pp. 33–33 (2005)
17. He, K., Lu, Y., Sclaroff, S.: Local descriptors optimized for average precision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
18. Hughes, G.B., Chraibi, M.: Calculating ellipse overlap areas. Computing and Visualization in Science **15**(5), 291–301 (2012)
19. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
20. Luo, Z., Shen, T., Zhou, L., Zhu, S., Zhang, R., Yao, Y., Fang, T., Quan, L.: Geodesc: Learning local descriptors by integrating geometry constraints. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
21. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(10), 1615–1630 (2005)
22. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision **65**(1-2), 43–72 (2005)
23. Miksik, O., Mikolajczyk, K.: Evaluation of local detectors and descriptors for fast feature matching. In: Proceedings of the International Conference on Pattern Recognition (ICPR). pp. 2681–2684 (2012)
24. Mishchuk, A., Mishkin, D., Radenovic, F., Matas, J.: Working hard to know your neighbor's margins: Local descriptor learning loss. In: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (NIPS). pp. 4829–4840 (2017)
25. Mishkin, D., Radenovic, F., Matas, J.: Repeatability is not enough: Learning discriminative affine regions via discriminability. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)

26. Moreels, P., Perona, P.: Evaluation of features detectors and descriptors based on 3D objects. International Journal of Computer Vision **73**(3), 263–284 (2007)
27. Mukundan, A., Tolias, G., Chum, O.: Multiple-kernel local-patch descriptor. In: British Machine Vision Conference (BMVC) (2017)
28. Ono, Y., Trulls, E., Fua, P., Yi, K.M.: Learning local features from images. In: Proceedings of the Conference on Neural Information Processing Systems (NIPS) (2018)
29. Pultar, M., Mishkin, D., Matas, J.: Leveraging outdoor webcams for local descriptor learning. In: Proceedings of the Computer Vision Winter Workshop (CVWW) (2019)
30. Schönberger, J.L., Hardmeier, H., Sattler, T., Pollefeys, M.: Comparative evaluation of hand-crafted and learned local features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017)
31. Simo-Serra, E., Trulls, E., Ferraz, L., Kokkinos, I., Fua, P., Moreno-Noguer, F.: Discriminative learning of deep convolutional feature point descriptors. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2015)
32. Snavely, N., Seitz, S., Szeliski, R.: Modeling the world from internet photo collections. International Journal of Computer Vision **80**(2), 189–210 (2008)
33. Strecha, C., von Hansen, W., Gool, L.J.V., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2008)
34. Tian, Y., Fan, B., Wu, F.: L2-Net: deep learning of discriminative patch descriptor in euclidean space. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6128–6136 (2017)
35. Tian, Y., Yu, X., Fan, B., Wu, F., Heijnen, H., Balntas, V.: SOSNet: Second order similarity regularization for local descriptor learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)
36. Wang, Z., Fan, B., Wang, G., Wu, F.: Exploring local and overall ordinal information for robust feature description. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(11), 2198–2211 (2016)
37. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 603–610 (2011)
38. Xu, Y., Gong, M., Liu, T., Batmanghelich, K., Wang, C.: Robust angular local descriptor learning. In: Proceedings of the Asian Conference on Computer Vision (ACCV) (2018)
39. Yi, K.M., Trulls, E., Lepetit, V., Fua, P.: LIFT: Learned invariant feature transform. In: Proceedings of the European Conference on Computer Vision (ECCV) (2016)
40. Yi, K., Verdie, Y., Fua, P., Lepetit, V.: Learning to assign orientations to feature points. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1–8 (2016)
41. Zitnick, C.L., Ramnath, K.: Edge foci interest points. In: IEEE International Conference on Computer Vision (ICCV). pp. 359–366 (2011)