Robust Feature-Based Image Matching Through Dual Descriptors

Javid Norouzi, Mohammad Sadegh Helfroush, Alireza Liaghat and Habibollah Danyali

February 1, 2024

Robust Feature-Based Image Matching Through Dual Descriptors

Javid Norouzi[1[0000−0002−3961−8823]], Mohammad Sadegh Helfroush[1[0000−0001−9095−4913]], Alireza Liaghat[1[0000−0002−3388−0871]], and Habibollah Danyali[1[0000−0002−3594−1674]]

Shiaz University of Technology, Shiraz, Iran {j.norouzi,ms_helfroush a.liaghat, danyali}@sutech.ac.ir

Abstract. Feature matching is a fundamental vision task but finding correct correspondences between images remains challenging. We propose a consensus-based dual descriptor matching approach to improve the reliability of feature correspondence between images. The key idea is to extract two distinct descriptors per keypoint and only keep matches that agree across both descriptor types. This provides an effective way to filter incorrect matches by requiring consistency between complementary descriptions. We implement and validate this dual descriptor strategy using classical hand-crafted features like SIFT, ORB, and BRISK, as well as learned descriptors from a deep network. Experiments on the HPatches benchmark demonstrate that our general approach of using dual descriptors consistently increases matching accuracy and enables matchability prediction, outperforming individual methods. The dual-descriptor consistency imparts robustness to variations in viewpoint and illumination conditions.

Keywords: Feature detection · Keypoint Descriptor · Dual descriptor.

1 Introduction

Feature matching between images is a fundamental task in computer vision with applications in object recognition, image stitching, 3D reconstruction, and more. Finding correct feature correspondences across images is challenging due to changes in viewpoint, illumination, clutter, and noise. Mismatched features can severely degrade subsequent tasks that rely on accurate matches.

Recent approaches have sought to improve feature matching accuracy and robustness. Some methods learn robust feature descriptors that are invariant to common image variations [9,3,18]. Other techniques use additional information like feature co-occurrence [19] or spatial verification [12] to find the correct matches. However, most methods match features independently using nearest-neighbor search on a single descriptor per keypoint.

In this work, we propose a feature matching strategy that computes two distinct descriptors per keypoint to improve matching accuracy. The key insight is that correct matches should be consistent across different descriptor types, while incorrect matches are less likely to persist. By only keeping feature matches that agree for both descriptor types, we obtain a more reliable set of correspondences. This simultaneously improves the accuracy of matches and reduces the number of outliers for subsequent processing stages like geometric model estimation.

An important consideration when matching image features is assessing the validity of the matches. Using two distinct descriptors per keypoint enables evaluating match quality in several ways. First, it allows determining if a pair of images contains enough consistent matches to conclude they show the same underlying scene. Second, it provides a way to judge if the extracted features and descriptors are robust enough to find reliable correspondences between the images.

We evaluate our approach using both classic hand-crafted descriptors like SIFT [3] and ORB [18], as well as learned descriptors from a deep network. Our two-descriptor matching strategy consistently improves results over single descriptors on standard benchmarks. We also show it is possible to train a network to output complementary descriptors that effectively filter matches when used together. The consistent improvement demonstrates the promise of multi-descriptor techniques for robust feature matching across challenging imaging conditions. Our key contributions include:

- Proposing a feature matching strategy that computes two distinct descriptors per keypoint to improve matching accuracy. Matching is done independently with each descriptor, and only matches that agree across both are kept.
- Demonstrating that using two complementary descriptors provides more reliable correspondences by reducing outliers. This simultaneously improves matching precision and reduces iterations needed for geometric model estimation.
- Training a deep network to output paired descriptors that are diverse and complementary. Applying the dual-descriptor strategy with learned features further improves matching accuracy.
- Discussing extensions like using the dual-descriptor agreement to determine if image pairs show the same underlying scene for recognition applications.

The rest of this paper is organized as follows. We review some of the related works in section 2. The proposed method is introduced in section 3. Experimental results are discussed in section 4 and the conclusion is drawn in section 5.

2 Related Works

Feature matching is a fundamental computer vision task critical for many applications like image retrieval, 3D reconstruction, and simultaneous localization and mapping (SLAM). Classic hand-crafted descriptors like SIFT [3], BRISK [10] and SURF [2] have been widely used, but exhibit limited robustness to variations in viewpoint, scale, and illumination. More recent hand-crafted feature detection and descriptions such as FAST-BRISK [13] and FDD [7] have tried to improve the feature description quality for more reliable feature matching.

Recent methods have sought to improve matching by learning robust keypoint descriptors from image data. SuperPoint [5] and D2-Net [6] learn detectors and descriptors end-to-end from data. HardNet [14] uses a deep network with triplet loss to learn invariance to transformations. More recent work like ALIKE [22] demonstrates improved keypoint detection through optimal assignment loss. ALIKED [20] further advances descriptor learning using deformable convolutions to achieve state-of-the-art performance. These learned descriptors outperform classic hand-crafted features by leveraging large training datasets. However, they can still be limited in generalizability by dependencies on the data distribution at training time.

Other works have proposed fusing or ensembling multiple descriptors to improve matching reliability. Norouzi et al. [16] extracted descriptors for the same keypoints over rotated versions of the image and concatenated them. Jun et al. [8] combined multiple descriptors from the same network to create a richer descriptor. Likewise, Norouzi et al. [17] used descriptors from different layers but instead of concatenating them, combined them using random projection. However, most prior ensembling approaches add computational load or show limited improvement in matching results.

Our work proposes improving feature matching by computing two distinct descriptors per keypoint. Requiring consensus between the different descriptor types provides a principled way to filter incorrect matches without adding any noticeable complexity. We demonstrate consistent gains over single descriptors on standard benchmarks.

3 Methodology

In this section, we introduce our proposed approach for improving feature matching accuracy using dual learned descriptors. Our approach consists of two different stages: 1) Feature detection and description (Using both classic algorithms and learned features) and 2) Feature matching. This approach requires two distinct feature extractors capable of producing uncorrelated descriptors. Uncorrelation is necessary to provide two independent judgments for matching keypoints, resulting in more correct matches. We present the methodology in three parts; first applying dual descriptor extraction using hand-crafted classic algorithms, implementing paired learned descriptors within an end-to-end deep network, and finally consensus-based matching.

3.1 Hand-Crafted Dual Descriptors

Figure 1 illustrates our scheme for extracting two sets of descriptors using classic algorithms. We first extract keypoints and their descriptors from the input image using "Feature Extractor 1". The same set of keypoints are then used to extract a second set of descriptors using "Feature Extractor 2". The main idea is to

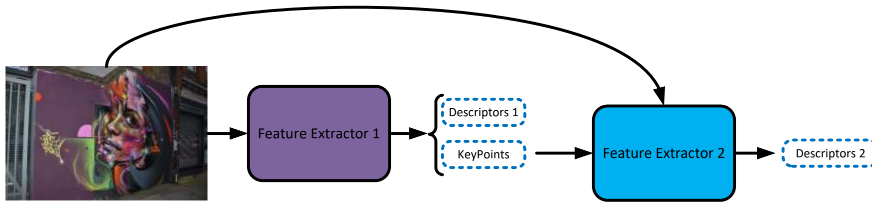


Fig. 1: The Proposed double descriptor strategy to match images using classic algorithms. A pair of descriptors are extracted for the same keypoints using two distinct Feature extraction algorithms.

compute a pair of descriptors for each keypoint. In practice, different feature extraction algorithms describe keypoints differently. Parameters of keypoints in one algorithm may not be suitable for another algorithm, preventing descriptor extraction for all keypoints. This is especially true about the keypoints scale. The scale of a keypoint defines the image area used to compute its descriptor. Therefore, it is necessary to tune the keypoint parameters so that both extractors can compute descriptors. This requires a knowledge of the inner workings of feature extractor algorithms. A straightforward approach is to assign arbitrary random descriptors for keypoints that are infeasible to describe based on their keypoint parameters.

3.2 Learned Dual Descriptors

To further demonstrate the benefits of our proposed dual descriptor approach, we also train a paired descriptor configuration for the ALIKED network. We modify the ALIKED architecture by adding a second descriptor head, consisting of additional convolutional layers paralleling the original. The second descriptor head has the same inner elements as the first descriptor. By freezing the original weights we train the second head to output complementary descriptors for each keypoint. Figure 2 shows our proposed architecture for our dual deascriptor ALIKED network.

ALIKED used MegaDepth [11] dataset for training. Since we did not have the resources to train on the full MegaDepth dataset, we took a simplified approach to learn our second descriptor. We generate synthetic image pairs with random affine transformations and use the known geometry to identify corresponding descriptors.

We train with the FastAP loss [4] from PyTorch Metric Learning [15]. To encourage distinct, uncorrelated descriptors, we provide original ALIKED descriptors along with the second descriptor to the loss. This enables the network to output a unique set of descriptors per keypoint.

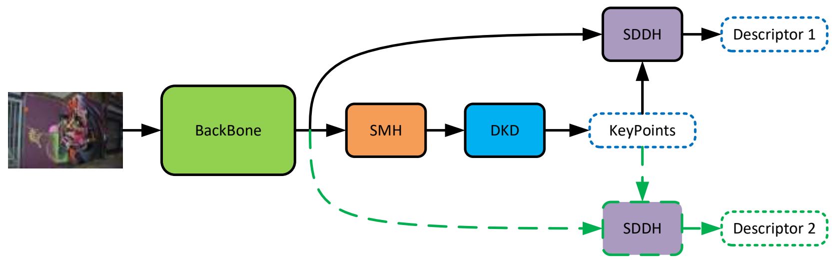


Fig. 2: Overview of our proposed learned dual descriptor extraction approach. We modify the ALIKED architecture by adding a second parallel descriptor head (dashed green).

Despite using less data, we demonstrate that our consensus matching with dual ALIKED descriptors still improves performance. In the next section, we provide experiments validating the benefits of our approach over the original single ALIKED descriptor.

3.3 Feature Matching

In the second stage(Figure 3) we will utilize the extracted keypoints and descriptor sets to assess the matchability between the images. Determining matchability is important, as it indicates whether the two images contain the same underlying scene. It also provides insight into whether the feature extraction algorithms can reliably match keypoints between the image pair. The "Common Match Finder" block determines the number of consensus matches across both descriptor types. Understanding matchability in advance allows us to gauge confidence in the final matching results.

A simple way to identify the consensus matches is to record the occurrence of each match in a combined matching matrix. If the number of keypoints in image 1 and image 2 are m and n respectively, the matching matrix $M_{[m,n]}$ for one set of descriptors can be computed as:

$$M[i,j] = \begin{cases} 1, & \text{if } P1_i \text{ and } P2_j \text{ are matched} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Where $P1$ and $P2$ are the keypoints from image 1 and image 2. We compute matching matrices M_1 and M_2 for both descriptor sets. The combined occurrence matrix $M_{\text{Occurrence}}$ is then as follows:

$$M_{\text{Occurrence}} = M_1 + M_2 \tag{2}$$

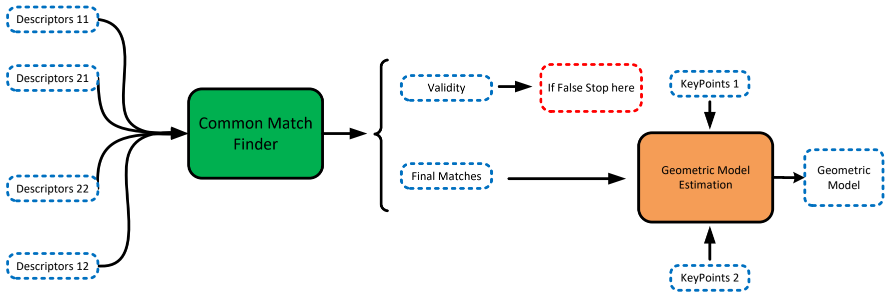


Fig. 3: Two sets of descriptors (Descriptors 11 and Descriptors 12 for the first image and Descriptors 21 and Descriptors 22 for the second image) are independently matched. The persistent matches are the final reliable matches. If the number of final matches is below a minimum threshold we stop proceeding and if not we continue estimating the geometric model using keypoints from both images.

Elements, where $M_{\text{Occurrence}}$ equals 2, indicate consensus matches found by both descriptor types.

If the number of common matches found is above a certain threshold, we can conclude the images are of the same scene with high probability. Low match numbers, on the other hand, imply the algorithms may struggle to establish correspondences between the images independently or the images don't include the same underlying scene. Overall, predicting matchability facilitates more robust matching and informs the appropriate interpretation of results. Algorithm 1 shows the overall procedure needed for this. The final matches alongside the image's keypoints are then used for geometric model estimation.

4 Experimental Results

In this section, we present experimental results and discussion to analyze the different components of our proposed approach. We evaluate the dual descriptor matching scheme and the matchability assessment separately to demonstrate their effectiveness on HPatches[1] dataset. Our experiments utilize combinations of classic feature extractors including SIFT, ORB, and BRISK as the dual descriptors as well as our dual ALIKED descriptor network.

Specifically, we first quantitatively compare the matching performance of individual versus dual descriptors on the HPatches dataset. This validates that combining multiple descriptor types improves accuracy and robustness compared to any single descriptor. It also enables geometric model estimation algorithms like RANSAC [21] to converge faster by reducing outliers.

Next, we assess the matchability prediction accuracy using the consensus-based descriptor agreement analysis on matched and mismatched HPatches im-

Descriptors11 D_{11}, Descriptors12 D_{12}, Descriptors21 D_{21}, Descriptors22 D_{22}, Minimum Matches $Min_{Matches}$ **Output:** Final Matches M_{Final}, and Matching Confirmation $Validity$

$Matches_1 \leftarrow$ Nearest neighbor search between D_{11} and D_{12} ;
$Matches_2 \leftarrow$ Nearest neighbor search between D_{21} and D_{22} ;
$M_{Final} \leftarrow$ Common matches between $Matches_1$ and $Matches_2$ (See equation 2);
$N_{Matches} \leftarrow$ Number of final common matches (M_{Final})
if $N_{Matches} > Min_{Matches}$ **then**
| $Validity \leftarrow True$
else
| $Validity \leftarrow False$
end

Algorithm 1: Matching Evaluation

age pairs. We show it reliably discriminates between pairs depicting the same versus different scenes prior to feature matching.

4.1 Dataset

To train the dual ALIKED descriptor network, we used an openly available dataset from Kaggle[1] containing approximately 82,000 random images spanning categories such as humans, portraits, animals, and more. We trained using the Adam optimizer with a learning rate of 10^{-3} for one epoch with a batch size of 256. The synthesized image pairs were sized at 392×392 pixels.

For evaluation, we utilize the HPatches Sequence dataset. This dataset consists of 116 image sets, where each set contains a reference image and 5 additional images depicting the same scene under varying viewpoints or illumination. The ground truth homographies relating the reference to other images are provided. The availability of ground truth transformations enables quantitative evaluation.

4.2 Evaluation metrics

We utilize the following metrics to quantitatively evaluate the performance of our proposed dual descriptor matching approach:

- N_{all} - The total number of initial matches found between the image pair.
- $NOCC$ - The number of geometrically consistent inlier matches after transformation estimation and outlier filtering. A higher inlier count indicates more accurate matching.
- $ROCC$ - The ratio of inliers to total matches. Higher values indicate fewer false matches.

$$\frac{NOCC}{N_{all}} \tag{3}$$

[1] https://www.kaggle.com/datasets/starktony45/image-dataset

- $RMSE$ -To evaluate the accuracy of the estimated transformation model f, we can compute the root mean squared error (RMSE) between points projected by f and the ground truth transformation g^{-1} as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N}\|g^{-1}(f(P_i)) - P_i\|^2} \tag{4}$$

  Where P_i are the keypoints in one image and N is the total number of keypoints. g^{-1} represents the inverse of the known homography relating to the image pair. This RMSE between mapped points measures how closely the estimated f aligns the images compared to the ground truth g.
- Precision - The fraction of identified same-scene pairs that are correct:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{5}$$

- Recall - The fraction of correct same-scene pairs that are identified:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{6}$$

- F1 score - The harmonic mean of precision and recall:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

- Accuracy - The fraction of predictions that are correct:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total population}} \tag{8}$$

4.3 Experimental Results and Discussion

In our experiments, we evaluate four variants of our proposed dual descriptor approach using combinations of SIFT, ORB, and BRISK for hand-crafted classic algorithms. We also evaluated our dual ALIKED Descriptor approach alongside them. The evaluated methods include:

- ORB (detection + description) + SIFT (description)
- ORB + BRISK
- BRISK + SIFT
- SIFT + BRISK
- Dual ALIKED descriptor

We assess five dual descriptor combinations on HPatches to validate their benefits over individual methods. Some image pairs lacked sufficient final matches for model estimation. These image pairs are analyzed separately. We also excluded 25 unstable BRISK pairs.

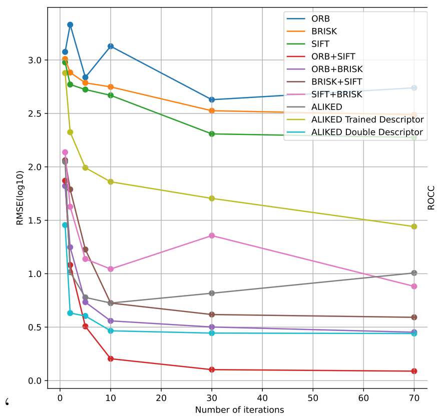
To analyze performance, we plot RMSE, ROCC, and inlier counts per RANSAC iteration (Figure 4). As outliers are filtered, inliers and alignment improve. The dual descriptors achieve lower RMSE (Figure 4a) and higher ROCC (Figure 4b) compared to individual methods, indicating more precise final alignment. However, the dual descriptors filter more correct matches, evidenced by lower inlier counts (Figure 4c).

We further analyzed pairs that lacked sufficient final matches (under 4 matches) in our dual descriptor approach. We used individual algorithms to match those image pairs in 70 RANSAC iterations. 16 pairs failed for BRISK+SIFT, 13 for SIFT+BRISK, and 72 pairs for ORB+BRISK while none of them succeeded individually. For ORB+SIFT, 82 pairs failed, but for 15 of the pairs either one or both of the individual approaches succeeded. For our dual ALIKED descriptor, 6 pairs failed. In all of these cases, the original ALIKED method succeeded, but our added second descriptor failed. This probably indicates that there is still room for learning better descriptors. Figure 4d shows the RMSE result for the failed pairs. Overall, our quantitative experiments validate that combining complementary descriptors improves matching robustness. The dual-descriptor agreement provides insight into difficult cases where current algorithms struggle to establish a reliable correspondence between keypoints.

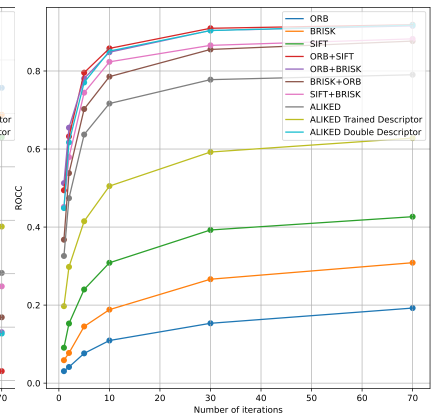
To further demonstrate the utility of our approach, we leverage the dual descriptors for scene recognition by predicting if two images show the same underlying location. Avoiding matching unrelated images saves computational effort. We generated 2048 random HPatches image pairs and recorded the consensus match counts using our method. By thresholding the counts, we set the minimum required to reliably identify related scenes. By trial and error, we have reported the best results here.

Table 1 presents the evaluation results using precision, recall, accuracy, and F1 score metrics. We excluded pairs with unstable results, mainly due to the high number of keypoints detected by BRISK, which would have required excessive resources to process. Overall, accuracy exceeded 90% for all methods. BRISK+SIFT achieved the highest accuracy and F1 score but required skipping approximately 600 unstable pairs. SIFT+BRISK attained the next best results with 0.976 accuracy and 0.966 F1 score using a minimum of 10 matches. While other combinations had lower minimum matches, SIFT+BRISK provided the best trade-off between performance and stability. In the hand-crafted algorithms leveraging dual descriptors provide an effective and efficient approach for scene recognition, validating the utility of our method beyond just feature matching.

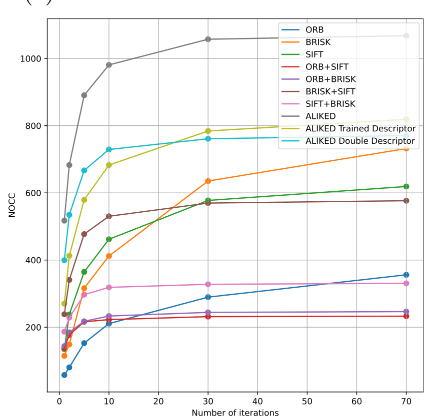
Interestingly, in the case of ALIKED, the number of matches was automatically very low for non-matching pairs. So there was no need for an extra descriptor for detecting when image pairs were not of the same scene. We simply applied a threshold on the original ALIKED matches in this case to classify pairs. ALIKED achieved an accuracy of 0.974 and F1 score of 0.974, competitive with the top dual hand-crafted methods.



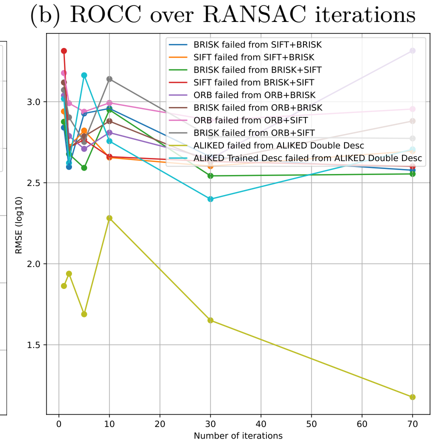
(a) RMSE over RANSAC iterations



(b) ROCC over RANSAC iterations



(c) Number of inliers over RANSAC iterations



(d) RMSE over RANSAC iterations for failed pairs using dual descriptors

Fig. 4: Simulation results including RMSE, ROCC, and NOCC using both dual descriptors and individual ones.

In terms of runtime, SIFT and BRISK are more expensive descriptors, so combining them incurs approximately the cost of both individually. For large images, this can be time-consuming. However, by precisely tuning the number of extracted keypoints, we can likely find an optimal trade-off between speed and performance.

In contrast, ORB is extremely efficient and detects just 500 keypoints by default. Computing additional descriptors has minimal overhead. In our experiments, we did not deeply optimize parameters for speed versus accuracy. But as a rule of thumb, 1k to 3k total keypoints provide a good balance for our method to function effectively. More keypoints incur wasted computation and resources.

Additionally, tuning the keypoint scales for descriptor extraction avoids unnecessary complexity.

For ALIKED the extra computational cost of adding a second descriptor is almost negligible. While we didn't use quantitative measures, based on hardware specifications available on Kaggle there was no noticeable difference in terms of performance compared to using ALIKED alone.

While we leave the extensive runtime analysis for future work, it is clear descriptor choice and keypoint settings provide levers for balancing matching quality versus efficiency. The dual extraction does introduce additional computation, but the cost can be controlled. For tasks demanding maximum robustness, the dual descriptors provide clear accuracy benefits.

Table 1: Scene recognition results comparing handcrafted and learning-based methods. Two descriptors were used for classical methods, but for the learning-based approach, one descriptor was sufficient.

Approach	Method	Min Matches	Precision	Recall	F1 Score	Accuracy
Handcrafted	ORB-SIFT	3	0.998	0.872	0.931	0.937
	ORB-BRISK	3	0.995	0.877	0.932	0.938
	BRISK-SIFT	4	0.998	0.965	0.981	0.998
	SIFT-BRISK	10	0.989	0.944	0.966	0.976
Learning Based	ALIKED	4	0.995	0.953	0.974	0.974

5 Conclusion

In this work, we presented a feature matching strategy that computes two distinct descriptors per keypoint to improve correspondence accuracy. Our key idea is to leverage consensus between different descriptor types to filter out false matches that are unlikely to persist across varying conditions. We evaluated our approach using combinations of classic hand-crafted descriptors. Across experiments on the HPatches benchmark, fusing multiple complementary descriptors consistently increased precision and recovered more geometrically consistent matches than individual techniques like SIFT, BRISK, ORB, or ALIKED. Our dual descriptor matching also enabled reliable matchability prediction, achieving over 90% accuracy in recognizing related images prior to geometric model estimation. In future work, we plan to explore combining more than two descriptors to further improve generalizability across imaging conditions. Optimizing the keypoint extraction and matching for efficiency is another important direction. Overall, our results demonstrate that consensus-based dual-descriptor matching provides a simple yet effective approach to enhance robustness in feature correspondence tasks. The technique can benefit applications like image retrieval, 3D reconstruction, and SLAM which rely on establishing accurate matches between challenging image pairs.

References

1. Balntas, V., Lenc, K., Vedaldi, A., Mikolajczyk, K.: Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. vol. 2017-January (2017). `https://doi.org/10.1109/CVPR.2017.410`
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer vision and image understanding **110**(3), 346–359 (2008)
3. Bellavia, F., Colombo, C.: Is there anything new to say about sift matching? International journal of computer vision **128**(7), 1847–1866 (2020)
4. Cakir, F., He, K., Xia, X., Kulis, B., Sclaroff, S.: Deep metric learning to rank. vol. 2019-June (2019). `https://doi.org/10.1109/CVPR.2019.00196`
5. Detone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. vol. 2018-June (2018). `https://doi.org/10.1109/CVPRW.2018.00060`
6. Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T.: D2-net: A trainable cnn for joint description and detection of local features. vol. 2019-June (2019). `https://doi.org/10.1109/CVPR.2019.00828`
7. Ghahremani, M., Liu, Y., Tiddeman, B.: Ffd: Fast feature detector. IEEE Transactions on Image Processing **30**, 1153–1168 (2021). `https://doi.org/10.1109/TIP.2020.3042057`
8. Jun, H., Ko, B., Kim, Y., Kim, I., Kim, J.: Combination of multiple global descriptors for image retrieval. arXiv preprint arXiv:1903.10663 (2019)
9. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: Binary robust invariant scalable keypoints. In: 2011 International Conference on Computer Vision. pp. 2548–2555 (2011). `https://doi.org/10.1109/ICCV.2011.6126542`
10. Leutenegger, S., Chli, M., Siegwart, R.Y.: Brisk: Binary robust invariant scalable keypoints (2011). `https://doi.org/10.1109/ICCV.2011.6126542`
11. Li, Z., Snavely, N.: Megadepth: Learning single-view depth prediction from internet photos (2018). `https://doi.org/10.1109/CVPR.2018.00218`
12. Liaghat, A., Helfroush, M.S., Norouzi, J., Danyali, H.: Airborne sar to optical image registration based on sar georeferencing and deep learning approach. IEEE Sensors Journal **23**(21), 26446–26458 (2023). `https://doi.org/10.1109/JSEN.2023.3314608`
13. Liu, Y., Zhang, H., Guo, H., Xiong, N.N.: A fast-brisk feature detector with depth information. Sensors (Switzerland) **18** (2018). `https://doi.org/10.3390/s18113908`
14. Mishchuk, A., Mishkin, D., Radenovi, F., Matas, J.: Working hard to know your neighbor's margins: Local descriptor learning loss. vol. 2017-December (2017)
15. Musgrave, K., Belongie, S.J., Lim, S.N.: Pytorch metric learning. ArXiv **abs/2008.09164** (2020)
16. Norouzi, J., Liaghat, A., Helfroush, M.S., Danyali, H.: A novel deep learning framework for image keypoint description. In: 2022 12th International Conference on Computer and Knowledge Engineering (ICCKE). pp. 332–336 (2022). `https://doi.org/10.1109/ICCKE57176.2022.9960057`
17. Norouzi, J., Liaghat, A., Helfroush, M.S., Danyali, H.: Vgg16-based feature fusion for image kyepoint description. In: 2023 31st International Conference on Electrical Engineering (ICEE). pp. 1005–1009 (2023). `https://doi.org/10.1109/ICEE59167.2023.10334668`
18. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: 2011 International conference on computer vision. pp. 2564–2571. Ieee (2011)

19. Xiao, G., Wang, S., Wang, H., Ma, J.: Mining consistent correspondences using co-occurrence statistics. Pattern Recognition **119**, 108062 (2021)
20. Xingming Wu, Weihai Chen, P.C.C.Q.X., Zhao, Z.L.X.: Aliked: A lighter keypoint and descriptor extraction network via deformable transformation. IEEE Transactions on Instrumentation & Measurement (2023)
21. Yaniv, Z.: Random sample consensus (ransac) algorithm, a generic implementation. The Insight Journal (2022). https://doi.org/10.54294/ia6mzx
22. Zhao, X., Wu, X., Miao, J., Chen, W., Chen, P.C., Li, Z.: Alike: Accurate and lightweight keypoint detection and descriptor extraction. IEEE Transactions on Multimedia (2022). https://doi.org/10.1109/TMM.2022.3155927