Supplementary Materials: RoSe: Rotation-Invariant Sequence-Aware Consensus for Robust Correspondence Pruning

Anonymous Authors

1 STATISTICS OF DATASETS

The initial correspondence set is obtained by SIFT provided in the VLFeat library. The correctness of each correspondence is determined by manually checking. We show the initial inlier ratios for the RS, UCD, and MR datasets in Figure 1. The average inlier ratios for these datasets are 68.50%, 25.37%, and 41.58%.

2 VISUALIZATIONS OF IMAGE MATCHING

We show the image matching results of our RoSe against other state-of-the-art methods on five representative image pairs: PAN90, CIAP129, UCD160, UCD169, and Retinal14. PAN90 is characterized by significant repeated structures, causing strong ambiguity in feature descriptors and leading to a low initial inlier ratio of 12.8%. CIAP129 features small overlapping areas, also resulting in a low initial inlier ratio of 10.1%. UCD160 and UCD169 involve rotations at varying scales. Retinal14, with a limited number of initial correspondences (156), presents a challenge for identifying consistency among inliers. As shown in Figure 2, RANSAC [2], LAF [3], and RFMSCAN [4] can identify most of the inliers with only a few misjudgments. LPM [7] is susceptible to a high ratio of outliers due to its unreliable neighborhood construction, resulting in some inliers being discarded. Both LOGO [8] and learning-based methods exhibit poor performance. The former relies on a set of high inliers for expanding the candidate set, but the high outlier ratios reduce the reliability of this set, resulting in unsatisfactory overall performance. Learning-based methods heavily depend on training data, exhibiting poor generalization ability on other types of data. As shown in Figure 3, for the image pair CIAP129, RANSAC exhibits poor performance, suggesting that random sampling struggles to find the optimal model within the given number of iterations under low inlier ratios. The remaining traditional methods perform comparatively better. Learning-based methods also fail to produce reasonable results, showing a significant number of misjudgments. As depicted in Figures 4 and 5 , the large-scale rotations in UCD160 and UCD169 render most methods ineffective. As shown in Figure 6, for Retinal14, LPM [7], and RFMSCAN [4] perform well, while other methods either discard a large number of inliers or exhibit more misjudgments. Our RoSe demonstrates good performance in all five image pairs, validating the effectiveness and generalization ability of the algorithm.

3 VISUALIZATIONS OF IMAGE REGISTRATION

We show the image registration results of our RoSe against other state-of-the-art methods on three image pairs, UAV1, SAR58, and CIAP120, as shown in Figures 7, 8 and 9. For UAV1, which involves projection transformation, methods such as LPM [7], LAF [3], and RoSe performed well, whereas other methods experienced registration distortion. RANSAC [2], with its inherent randomness, failed to consistently produce reliable registration results. Both LPM [7]

and LAF [3] demonstrated effective overall registration, aligning most areas adequately. However, due to unreliable neighborhood consistency, some local regions do not maintain inliers, thus causing distorted registration results. SAR58 features severe noise and low texture. For most methods, they exhibit noticeable distortions in the registration results except LAF [3] and RoSe. Most traditional methods successfully recover the valid transformation function. By contrast, many learning-based methods yield strange registration results, primarily due to their limited generalization capability, which is induced by the imposition of the essential matrix loss. CIAP120, with small overlapping areas and a high proportion of outliers, only showed favorable results with LPM [7], RFMSCAN [4], and RoSe. Since the rectified local neighborhood construction significantly enhances the reliability of neighborhoods for correspondence and aids in the rotation-invariant sequence-aware consistency, our Rose can identify most of inliers, thereby recovering accurate transformation function and producing high-quality registration results.

REFERENCES

[1] Luanyuan Dai, Yizhang Liu, Jiayi Ma, Lifang Wei, Taotao Lai, Changcai Yang, and Riqing Chen. 2022. MS2DG-Net: Progressive correspondence learning via multiple sparse semantics dynamic graph. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8973–8982.
[2] Martin A Fischler and Robert C Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communicaions of the ACM* 24, 6 (1981), 381–395.
[3] Xingyu Jiang, Jiayi Ma, Aoxiang Fan, Haiping Xu, Geng Lin, Tao Lu, and Xin Tian. 2021. Robust feature matching for remote sensing image registration via linear adaptive filtering. *IEEE Transactions on Geoscience and Remote Sensing* 59, 2 (Feb. 2021), 1577–1591.
[4] Xingyu Jiang, Jiayi Ma, Junjun Jiang, and Xiaojie Guo. 2020. Robust feature matching using spatial clustering with heavy outliers. *IEEE Transactions Image Processing* 29 (2020), 736–746.
[5] Xin Liu, Guobao Xiao, Riqing Chen, and Jiayi Ma. 2023. Pgfnet: Preference-guided filtering network for two-view correspondence learning. *IEEE Transactions on Image Processing* 32 (2023), 1367–1378.
[6] Xin Liu and Jufeng Yang. 2023. Progressive neighbor consistency mining for correspondence pruning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9527–9537.
[7] Jiayi Ma, Ji Zhao, Junjun Jiang, Huabing Zhou, and Xiaojie Guo. 2019. Locality preserving matching. *International Journal of Computer Vision* 127, 5 (May 2019), 512–531.
[8] Yifan Xia and Jiayi Ma. 2022. Locality-guided global-preserving optimization for robust feature matching. *IEEE Transactions on Image Processing* 31 (2022), 5093–5108.
[9] Shihua Zhang and Jiayi Ma. 2023. Convmatch: Rethinking network design for two-view correspondence learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
[10] Chen Zhao, Yixiao Ge, Feng Zhu, Rui Zhao, Hongsheng Li, and Mathieu Salzmann. 2021. Progressive correspondence pruning by consensus learning. In *Proceedings of the IEEE International Conference on Computer Vision*. 6464–6473.
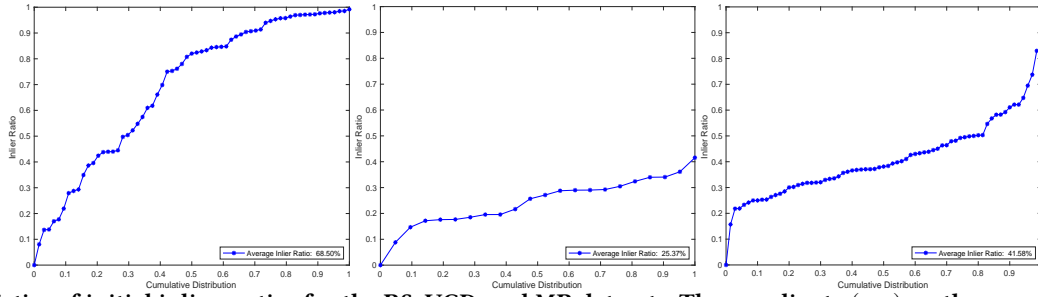
Figure 1: Statistics of initial inliers ratios for the RS, UCD, and MR datasets. The coordinate (x, y) on the curve indicates that there are $(100^*x)\%$ image pairs with initial inlier ratio not exceeding y.
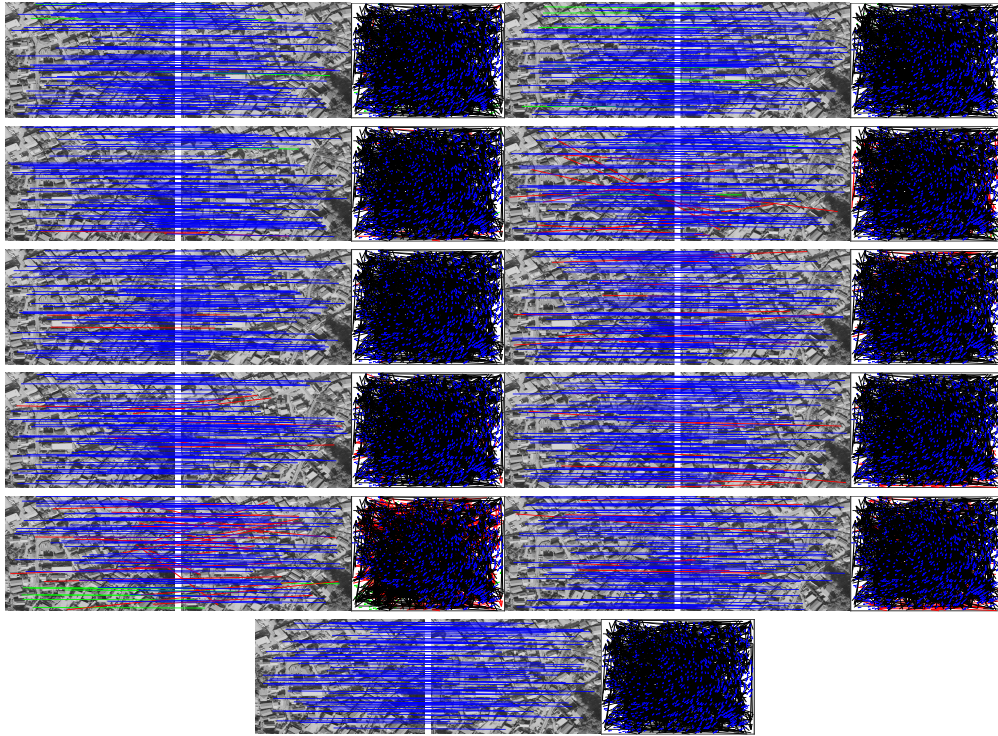

Figure 2: The visualization comparison of RANSAC, LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the PAN90 image pair for image matching (from top to bottom, left to right). Each group includes the image matching results along with their corresponding motion vector fields. The head and tail of each arrow in the motion vector field indicate the positions of the feature points in the two images (blue = true positive, black = true negative, green = false negative, red = false positive). For clarity, we randomly select up to 100 correspondences to display in each image pair, true negatives are not shown.
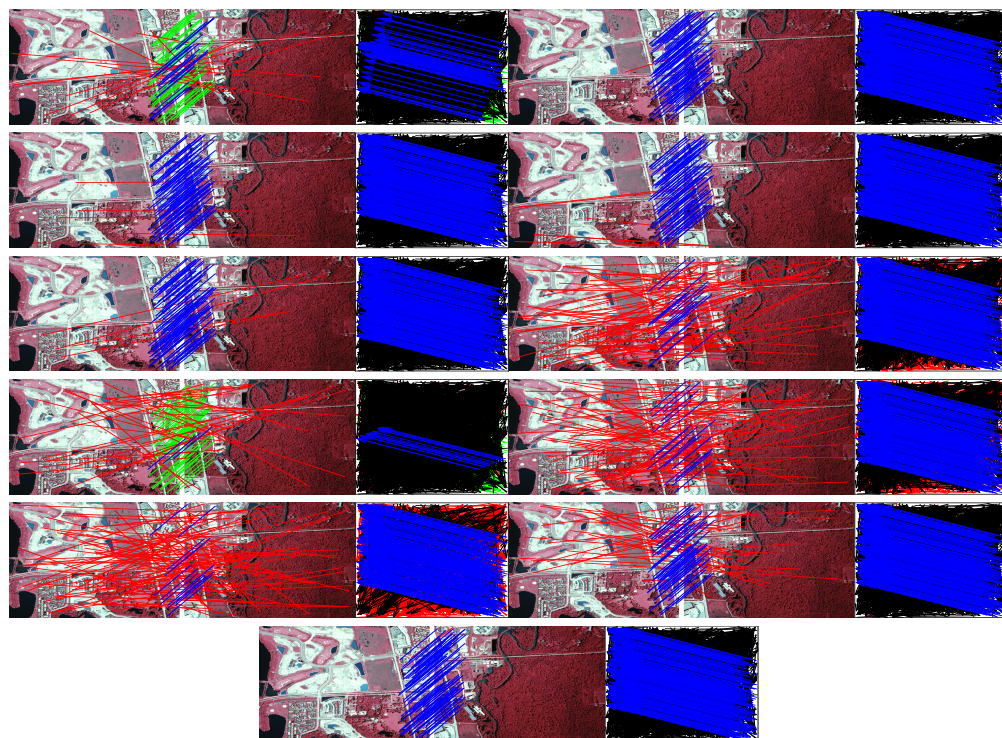
Figure 3: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the CIAP129 image pair.
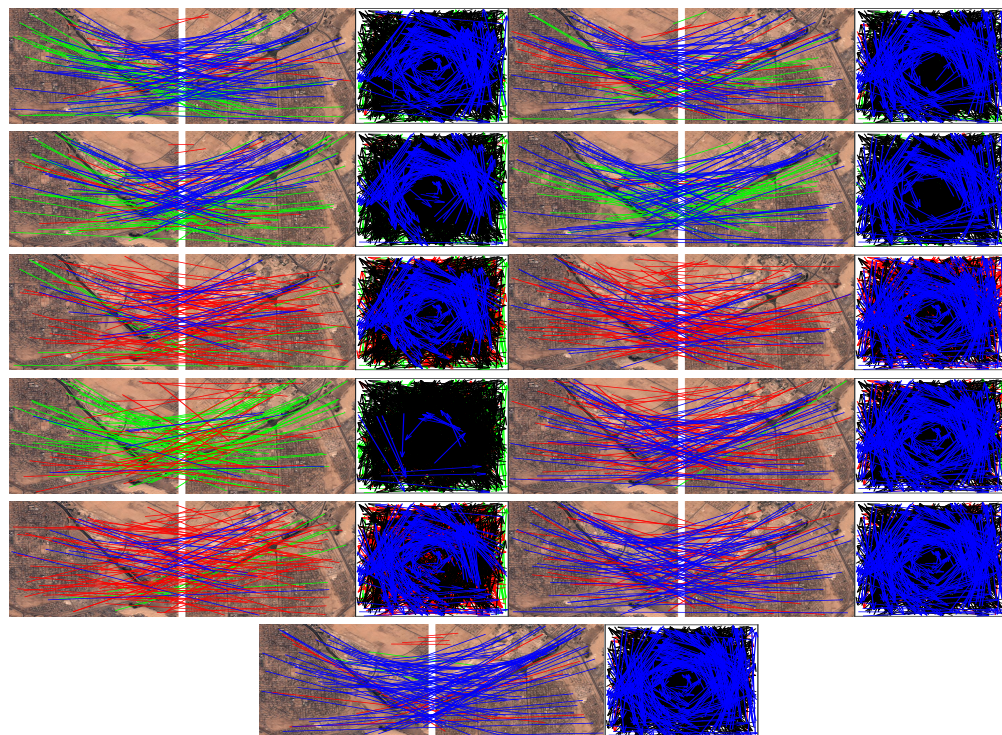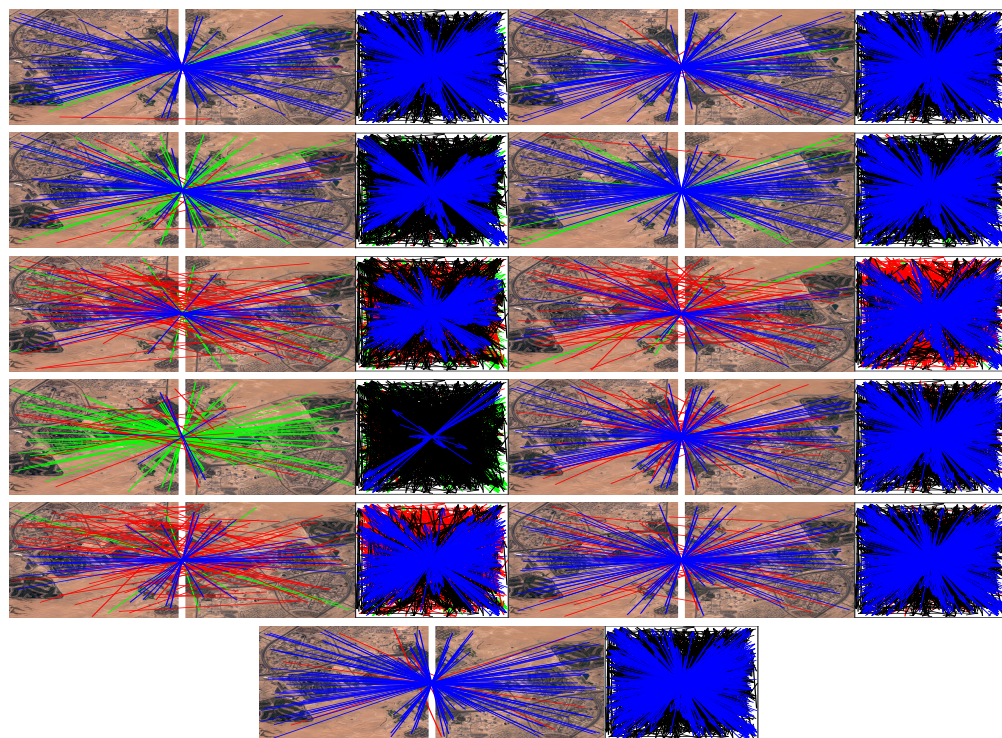

Figure 4: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the UCD160 image pair.

Figure 5: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the UCD169 image pair for image matching.
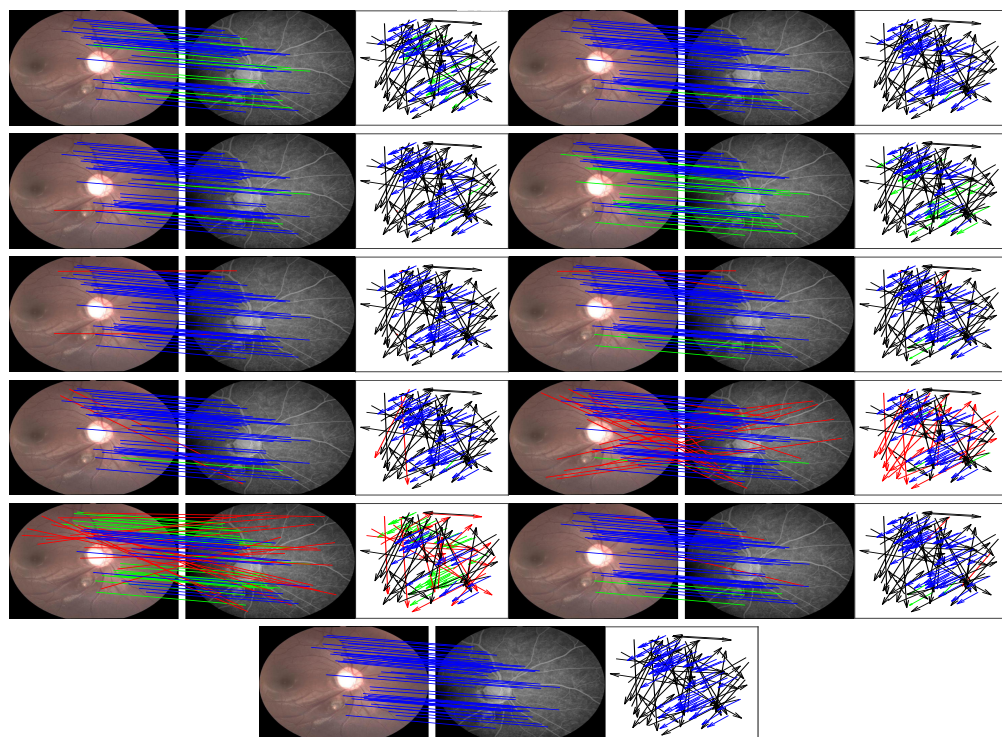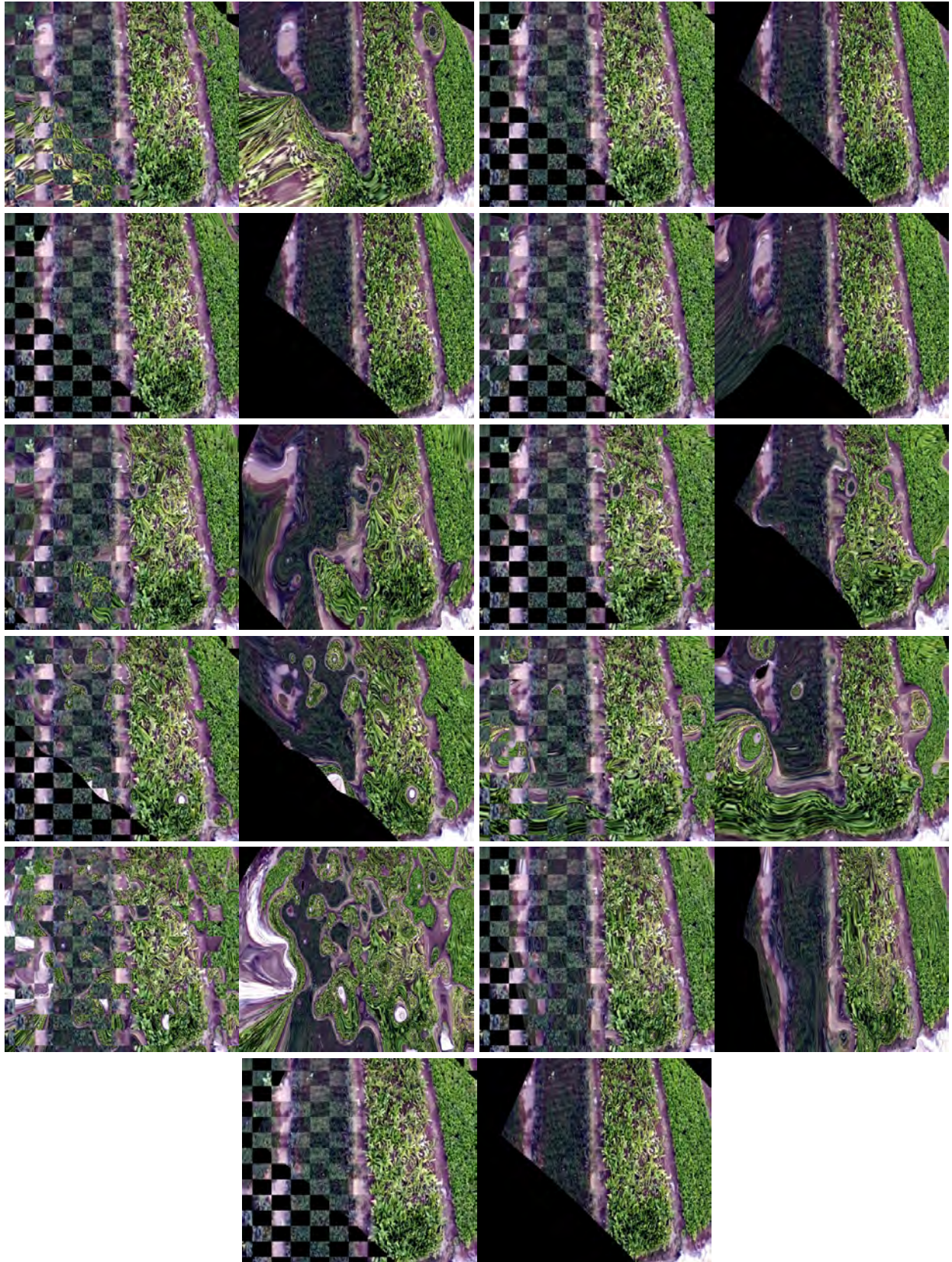

Figure 6: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the Retial14 image pair for image matching.

Figure 7: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the UAV1 image pair for image registration (from top to bottom, left to right).
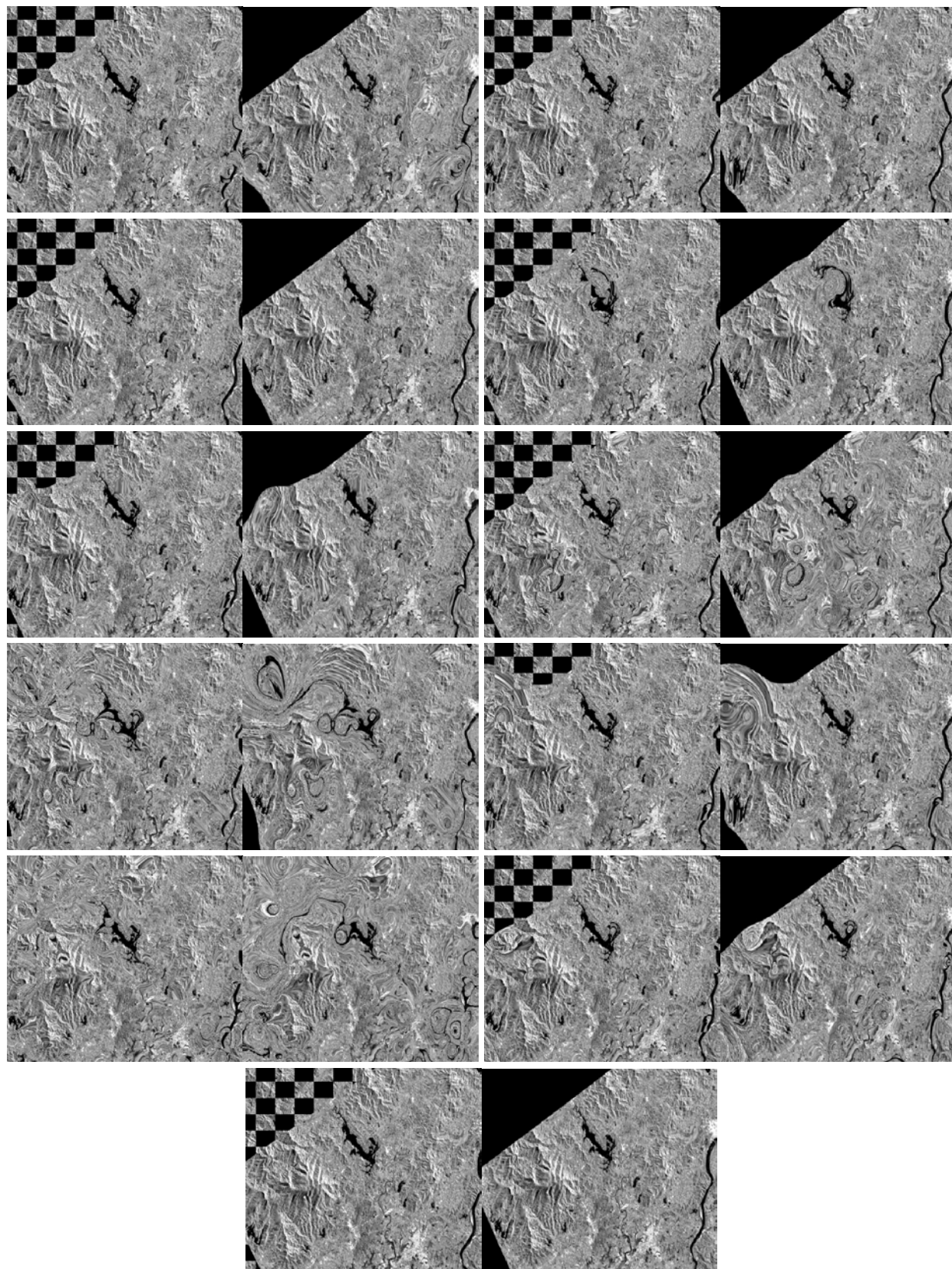
Figure 8: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the SAR58 image pair for image registration.
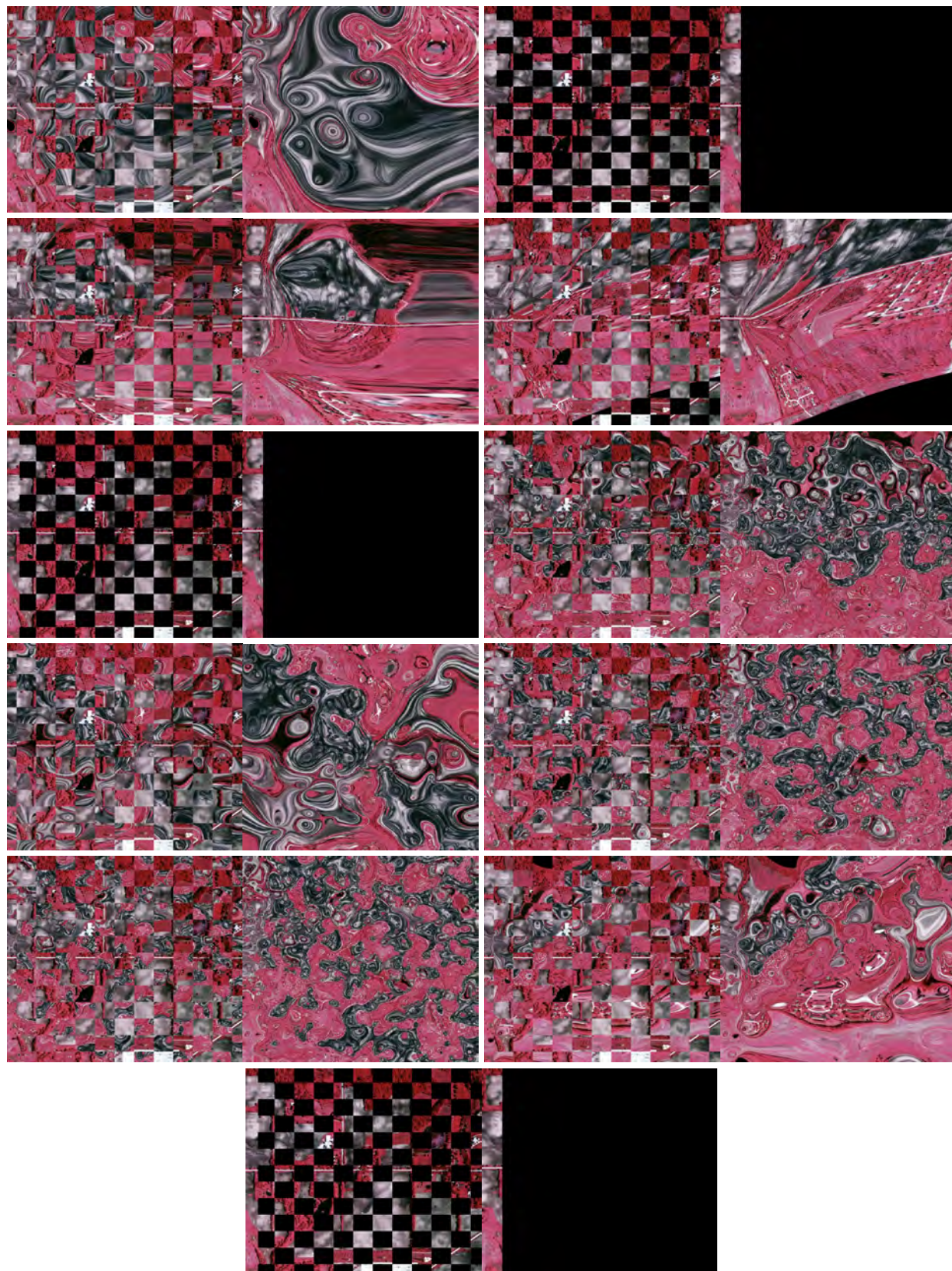
Figure 9: The visualization comparison of RANSAC [2], LPM [7], LAF [3], LOGO [8], RFMSCAN [4], MS2DGNet [1], NCMNet [6], PGFNet [5], ConvMatch [9], CLNet [10] and RoSe on the CIAP120 image pair for image registration.