Image matching via progressive priors

Weiqing Wang,[1,2] Yongrong Sun,[1] Kedong Zhao,[1]
Zhong Liu,[2] Wenjun Luo,[2] and Jinchang Qin[2]

[1]*College of Automation Engineering, Navigation Research Center, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China*

[2]*School of Mechanical and Electrical Engineering, Guilin University of Aerospace Technology, Guilin, 541004, China*
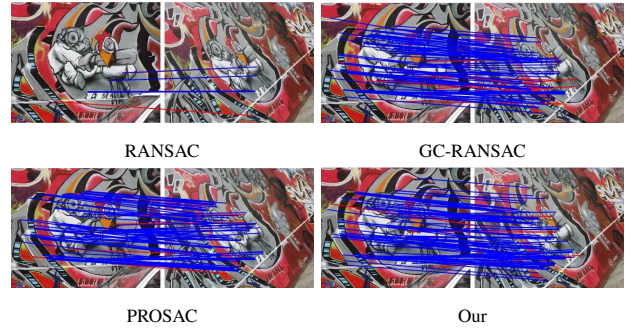Email: wangwq@guat.edu.cn

Fig 1 *Image matching results of RANSAC, GC-RANSAC, PROSAC, and our proposed method on the graffiti image pairs. The blue line signifies a match of true positive and the red line signifies a match of false positive. For visibility, at most 100 randomly selected matches are presented in the image pairs, and the false negatives are not shown.*

It is a challenging issue how to improve the accuracy of image matching in computer vision. To address this issue, an image matching method is proposed, which is via progressive priors of a putative dataset. Distance ratio priors of a putative dataset are initially employed to calculate a tentative deformation through geometric constraints. Progressive priors of the putative dataset, obtained by the tentative deformation, are then engaged to improve the accuracy of image matching by estimating a global deformation. The comparison experiments illustrate that our proposed method more effectively enhances the accuracy of image matching than six state-of-the-art methods.

Introduction: A fundamental issue in computer vision (e.g., image registration[1]) is establishing reliable correspondences in an image pair of the same scene. A two-stage strategy is usually used for image matching. A putative dataset is created first, and then mismatches are eliminated from them. Usually, a putative dataset is generated by some hand-crafted features with their similar feature descriptors (e.g., scale-invariant feature transform (SIFT[2]), oriented Fast and Rotated BRIEF[3] (ORB[4])). Because of the ambiguousness of the local descriptors, there are numerous false matches in the putative dataset. Therefore, how to remove mismatches from the putative dataset in an image pair has become a subject worth studying.

Random sample consensus (RANSAC[5]) is most popularly signed up to fit a parametric model for removing mismatches. Based on the assumption that inliers will tend to be closer to one another than outliers, Myatt et al.[6] propose N-adjacent points sample consensus (NAPSAC). Progressive sample consensus (PROSAC[7]) takes into account distance ratio priors of a putative dataset to estimate a global deformation. When a current best deformation is found, Graph-Cut RANSAC (GC-RANSAC[8]) uses the spatial coherence in the local optimization step to perform the graph cut approach. Many priors are fused to the estimation by Xiang et al.[9], such as distance ratio, orientation difference, and scale ratio. Ma et al.[10] propose pixel shift clustering RANSAC (PSC-RANSAC) to remove the residual false matches in the matching results of standard RANSAC.

In this letter, we propose an image matching method via progressive priors of a putative dataset. Distance ratio priors of a putative dataset are initially adopted to calculate a tentative deformation through geometric constraints. Progressive priors of the putative dataset, acquired by the tentative deformation, are then utilized to remove mismatches by estimating a global deformation. The main contribution of this work is that progressive priors of a putative dataset can be acquired by a tentative deformation, which is estimated through geometric constraints.

The remainder of this paper is organized as follows. Sections 2, 3, 4, and 5 describe the image matching method via progressive priors of a putative dataset in detail. Section 6 presents the experimental results to validate our proposed method. Section 7 provides a summary and concluding remarks.

distance ratio priors: Putative correspondences are generated by some hand-crafted features with similar feature descriptors. Suppose that N putative correspondences $C = \{c_i\}_{i=1}^N = \left\{(p_i, p'_i)\right\}_{i=1}^N$ are achieved based on the similarity of ORB feature descriptors in an image pair, where p_i and p'_i are the position of the i-th correspondence's two feature points in the image pair. w_{DR} represents the distance ratio. Thus

$$w_{DR} = \frac{dist_{nearest}}{dist_{sec-nearest}} \qquad (1)$$

where $dist_{nearest}$ and $dist_{sec-nearest}$ are the hamming distance between an ORB feature point descriptor and its nearest or second nearest neighbor in the second image. The smaller the distance ratio of correspondence, the more likely the correspondence is to be a true match. Therefore, distance ratio priors of a putative dataset are usually adopted to estimate the deformation of two images in PROSAC.

progressive priors: In image matching, the fundamental matrix, essential matrix, and homography matrix are restrictions between a pair of matching points from different views of the same scene. For convenience, this letter focuses on the homography matrix. Suppose that $p(x, y)$ and $p'(x', y')$ are two points of correspondence in image matching. According to the pinhole camera model, the point p is mapped from one perspective to the point p' in another perspective by a homography matrix $\mathbf{H}$, as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = H \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (2)$$

where $\mathbf{H}$ is a 3×3 matrix. $\mathbf{H}$ is usually normalized by scaling $h_{33} = 1$, and as such $\mathbf{H}$ can be parameterized by eight parameters. Typically, a homography matrix can be computed from four correspondences in an image pair. $p^r(x^r, y^r)$ is a reprojected feature point of $p(x, y)$ under the geometric model $\mathbf{H}$. Thus, w_{RE} represents the reprojection error.

$$w_{RE} = r(p, p') = d(p^r, p') \qquad (3)$$

Where $r(p, p')$ indicates the reprojection error, $d(p^r, p')$ indicates the Euclidean distance between p^r and p' points. When the reprojection error of correspondence is less than a given threshold, the correspondence could be assumed a true match. Therefore, the smaller the reprojection error of correspondence, the more possible the correspondence is to be a true match.

progressive sample consensus: Progressive sample consensus is an improvement of RANSAC, which takes into account the probability of the sample. A quality function e is initially utilized to sort the tentative correspondences in descending order.

$$c_i, c_j \in C : i < j \Rightarrow e(c_i) \geq e(c_j) \qquad (4)$$

The quality function of a sample m is defined as the lowest quality of a correspondence included in the sample.

$$e(m) = \min_{c_i \in C_m} e(c_i) \qquad (5)$$

Following the completion of the samples of size m out of N tentative correspondences, the samples are sorted in descending order using Eq. (5). Let T_q be an average number of samples from the tentative dataset which is a set of q correspondences.

$$T_q = T_N \frac{\binom{q}{m}}{\binom{N}{m}} = T_N \prod_{i=0}^{m-1} \frac{q-i}{N-i} \qquad (6)$$

Where T_N indicates the average number of samples from the whole dataset. According to Eq. (6), T_{q+1} could be recursively defined as

$$T_{q+1} = T_N \prod_{i=0}^{m-1} \frac{q+1-i}{N-i} = \frac{q+1}{q+1-m} T_q \qquad (7)$$

When $q < m$, $T_q = 0$, and with $T_N = 1$. When $q > m$, the value of T_q can be calculated through the Eq. (7). If the value of T_q is not an integer, $T'_q = 1$ is defined, and

$$T'_{q+1} = T'_q + \lceil T_{q+1} - T_q \rceil \qquad (8)$$

Therefore, the growth function is defined as

$$g(k) = \min \{ q : T'_q \geq k \} \qquad (9)$$

where k is the k-th sample. The sampling subset can be expanded according to Eq. (9). The iteration will stop when the probability that the correspondences are by chance inliers to an arbitrary wrong model is less than η. Furthermore, when the number of iterations reaches its maximum, the procedure will come to an end. In the process of estimating a homographic matrix, the number of iterations complies with the following constraints.

$$(1 - \beta^4)^k < 1 - \eta \qquad (10)$$

Where β indicates the inlier ratio. According to the Eq. (10),

$$k > \frac{\log(1-\eta)}{\log(1-\beta^4)} \qquad (11)$$

When k fulfills Eq. (11), it is supposed that an all-inliers sample has already occurred.

image matching via progressive priors: To eliminate false matches in a putative dataset, an image matching method is proposed, which is via progressive priors of a putative dataset. Our proposed method firstly adopts distance ratio priors of correspondences to estimate a tentative deformation of an image pair through geometry constraints. Progressive priors (e.g., reprojection error) are next achieved by the tentative homography matrix. The progressive priors of correspondences are then signed up to calculate a global deformation for improving the accuracy of image matching. The proposed method is an image matching method via progressive priors of a putative dataset, outlined in Algorithm 1.

Algorithm 1 image matching via progressive priors

1: **Input:** an image pair $\{I_1, I_2\}$, α, η, N, θ
2: **Output:** inliers
3: Detect ORB feature points and calculate their descriptors;
4: Generate a putative set $C = \{c_i\}_{i=1}^N$ using brute-force matching;
5: Sort the putative set $C = \{c_i\}_{i=1}^N$ in descending order based on $\frac{1}{w_{DR}}$ using Eq. (1);
6: Sample, computer a tentative homography matrix, and verify model using Eq. (8) and Eq. (11);
7: Calculate the reprojection error of the putative dataset using Eq. (2) and Eq. (3);
8: Sort the putative set $C = \{c_i\}_{i=1}^N$ in descending order based on $\frac{1}{w_{RE}}$ using Eq. (3);
9: Sample, computer a global homography matrix, and verify model using Eq. (8) and Eq. (11);
10: Identify inliers by the threshold θ.

Experimental results: To comprehensively evaluate the performance of our proposed method, we verify its performance with RANSAC, PROSAC, GC-RANSAC, GMS[11], LPM[12], and RFM-SCAN[13] on image pairs from two public datasets (e.g., VGG[14] and Heinly[15]). Figure 2 shows several examples from the VGG and Heinly datasets. These image pairs contain various changes, such as illumination, light, zoom, rotation, blur, and viewpoint. The two public datasets supply the ground truth.

To evaluate the performance of the image matching approach, the following metrics are used: precision, recall, and f-score.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (12)$$

$$\text{F} - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

(a) Trees   (b) Graffiti   (c) Boat
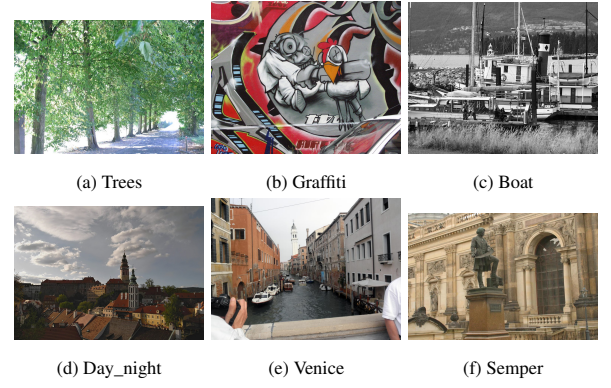
(d) Day_night   (e) Venice   (f) Semper

Fig 2 *Examples of image pairs in the VGG and Heinly datasets.*

Where TP stands for true positive, FP for false positive, and FN for false negative.

We compare our proposed method with RANSAC, PROSAC, GC-RANSAC, GMS, LPM, and RFM-SCAN on the image pairs. For fair comparisons, we use open-source implementations of these methods and set the parameters of each algorithm to the authors' specifications. LPM and RFM-SCAN are carried out in Matlab R2016a. Our proposed method, RANSAC, PROSAC, GC-RANSAC, and GMS are c++ code executing in Ubuntu 16.04. 10,000 ORB feature points are extracted in experiments through the open-source toolbox OpenCV 3.3. A putative dataset is generated by brute-force matching. The termination probability is set to 0.01 in RANSAC, PROSAC, GC-RANSAC, and our proposed method. The number of iterations is set to 500. In GMS, both scale and rotation functions are turned on. The reprojection threshold is set to 3 pixels in the VGG dataset or 2.5 pixels in Heinly Dataset to identify the true match. All experiments are implemented on a laptop PC with an Intel Core i5-7200U, 2.5 GHz CPU, and 16 GB of RAM.

We first provide intuitive results of our proposed method on the graffiti image pair as presented in Fig. 1, and the day_night image pair in Fig. 3. The interfering type of the graffiti and day_night image pair is viewpoint change, and illumination, respectively. Hence, it is a relative challenge to establish reliable feature correspondences. Figure 1 and Fig. 3 show that there are a small number of matches of true positive and a large number of matches of false negative in RANSAC. Because the inlier ratio of the graffiti and day_night image pair is 6.05%, and 3.45%, separately, RANSAC gets the worst performance. The f-score of RANSAC, GC-RANSAC, PROSAC, our proposed method in Fig.1 is 8.99%, 87.41%, 89.33%, and 95.14%, respectively. The f-score of RANSAC, GC-RANSAC, PROSAC, our proposed method in Fig.3 is 50.62%, 97.08%, 95.68%, and 98.69%, separately. We can notice that our proposed method achieves the best performance.
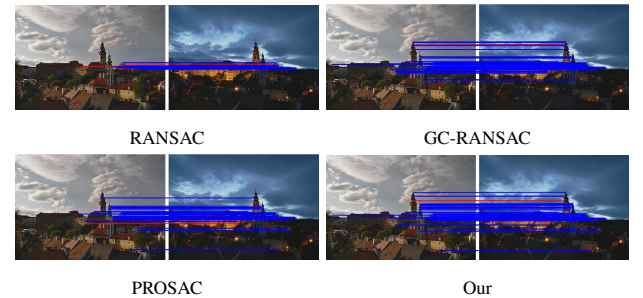

RANSAC   GC-RANSAC

PROSAC   Our

Fig 3 *Image matching results of RANSAC, GC-RANSAC, PROSAC, and our proposed method on the day_night image pairs. The blue line signifies a match of true positive and the red line signifies a match of false positive. For visibility, at most 100 randomly selected matches are presented in the image pairs, and the false negatives are not shown.*

Figure 4 shows the curves of inlier ratio, precision, recall, and f-score with respect to the cumulative distribution, which are the image matching results of the seven methods on the image pairs. When the inlier ratio

Table 1. Average running time of the six methods on the Oxford dataset. Bold indicates the best result.

	RANSAC	PROSAC	GC-RANSAC	GMS	LPM	RFM-SCAN	Our
Time (s)	0.021	**0.013**	0.310	0.252	0.163	15.752	0.025

is lower than 3%, RANSAC and GC-RANSAC fail. For fairly, we do not report these situations. In Fig. 4, the average inlier ratio of the image pairs is 32.66%. PROSAC has the best average precision (87.09%), followed by our proposed method (87.05%) and GC-RANSAC (86.86%). Because RANSAC is sensitive to the inlier ratio, its performance is hindered when the inlier ratio is low. RFM-SCAN obtains the best average recall (97.25%), followed by our proposed method (97.04%) and GC-RANSAC (96.70%). Except for LPM, the other six methods achieve a similar performance when the inlier ratio is higher than 27.03%. Our proposed method obtains the best average f-score (90.73%), followed by GC-RANSAC (90.46%) and PROSAC (90.44%). When the inlier ratio is higher than 45%, RANSAC achieves similar performance to our proposed method, GC-RANSAC, and PROSAC. Figure 4 demonstrates that our proposed method has a stronger performance to establish reliable correspondences than RANSAC, PROSAC, GC-RANSAC, GMS, LPM, and RFM-SCAN.
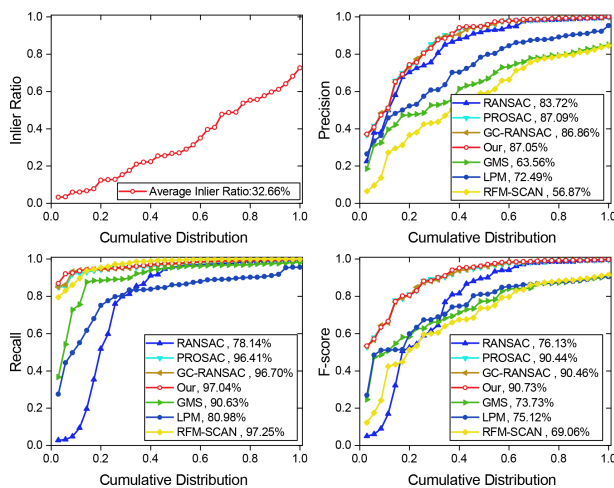

Fig 4 *Inlier ratio (top, left), precision (top, right), recall (bottom, left), and f-score (bottom, right) of RANSAC, PROSAC, GC-RANSAC, GMS, LPM, RFM-SCAN, and our proposed method with respect to the cumulative distribution on the image pairs from the VGG and Heinly datasets. The numbers in the boxes represent the average inlier ratio, precision, recall, and f-score.*

Table 1 provides the average running time (excluding the cost of ORB feature extraction and brute-force matching) results. RANSAC, PROSAC, GC-RANSAC, GMS, LPM, RFM-SCAN, and our proposed method, acquires 0.021, 0.013, 0.310, 0.252, 0.163, 15.752, 0.025 seconds, separately. Table 1 shows that PROSAC takes the least time compared to RANSAC, GC-RANSAC, and our proposed method. The running time of our proposed method is similar to the RANSAC. It illustrates that the running time of our proposed method is more than doubles that of PROSAC for better performance.

Conclusion: This letter reports an image matching method via progressive priors. The ability to obtain progressive priors of a putative dataset using a tentative deformation estimated by geometric constraints is a crucial feature of our method. Distance ratio priors of a putative dataset are firstly employed to estimate a tentative deformation through geometric constraints. Progressive priors of the putative dataset, acquired by the tentative deformation, are then adopted to improve the accuracy of image matching by calculating a global deformation. Our proposed method outperforms RANSAC, PROSAC, GC-RANSAC, GMS, LPM, and RFM-SCAN in comparative experiments using the image pairs from the VGG and Heinly datasets.

References

1. Zeng, L., et al.: Analysis of feasibility and advantages of multi-source image registration using regional features. Electronics Letters 57(7), 285–287 (2021)

2. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision 60(2), 91–110 (2004). [doi:10.1023/b:visi.0000029664.99615.94]

3. Calonder, M., et al.: Brief: Binary robust independent elementary features. In: European conference on computer vision, pp. 778–792. Springer (2010). [doi:10.1007/978-3-642-15561-1_56]

4. Rublee, E., et al.: Orb: An efficient alternative to sift or surf. In: 2011 International conference on computer vision, pp. 2564–2571. Ieee (2011). [doi:10.1109/iccv.2011.6126544]

5. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6), 381–395 (1981). [doi:10.1016/b978-0-08-051581-6.50070-2]

6. Torr, P.H., Nasuto, S.J., Bishop, J.M.: Napsac: High noise, high dimensional robust estimation-it's in the bag. In: British Machine Vision Conference (BMVC), (2002)

7. Chum, O., Matas, J.: Matching with prosac-progressive sample consensus. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1, pp. 220–226. IEEE (2005). [doi:10.1109/cvpr.2005.221]

8. Barath, D., Matas, J.: Graph-cut ransac. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6733–6741. (2018). [doi:10.1109/cvpr.2018.00704]

9. Xiang, H., et al.: Matching with guisac-guided sample consensus. IEICE TRANSACTIONS on Information and Systems 104(2), 346–349 (2021)

10. Ma, S., et al.: An image matching optimization algorithm based on pixel shift clustering ransac. Information Sciences 562, 452–474 (2021)

11. Bian, J., et al.: Gms: Grid-based motion statistics for fast, ultra-robust feature correspondence. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4181–4190. (2017). [doi:10.1109/cvpr.2017.302]

12. Ma, J., et al.: Locality preserving matching. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 4492–4498. (2017). [doi:10.24963/ijcai.2017/627]

13. Jiang, X., et al.: Robust feature matching using spatial clustering with heavy outliers. IEEE Transactions on Image Processing 29, 736–746 (2019). [doi:10.1109/tip.2019.2934572]

14. Mikolajczyk, K., et al.: A comparison of affine region detectors. International journal of computer vision 65(1), 43–72 (2005). [doi:10.1007/s11263-005-3848-x]

15. Heinly, J., Dunn, E., Frahm, J.M.: Comparative evaluation of binary features. In: European Conference on Computer Vision, pp. 759–773. Springer (2012)