Comparing Strategies of Random Sample Consensus Algorithms for Small Unmanned Aerial Vehicles Using Structure From Motion

Yeana L Bond, Esmeralda Osornio, Spencer Ledwell, Alberto C Cruz

California State University, Bakersfield, USA

Abstract

Unmanned Aerial Vehicles (UAVs) are used in diverse ways. The focus of our work is application of Structure from Motion (SFM) to surface identification, an important task for UAVs. SFM algorithms employ Random Sample Consensus (RANSAC) in several steps of their pipeline for recognition and remote sensing. We examine the evolution of the RANSAC algorithm and its various implementations. Through comparison, we learn that improving accuracy is a chief concern, and the varied ways RANSAC algorithms accomplish this goal. We discuss challenges encountered when comparing RANSAC variants and share our insight that in addition to accuracy, performance and robustness should be considered to equip small, task-oriented UAVs using a pipeline of SFM with a cost-efficient estimator.

1. Introduction

Use of Unmanned Aerial Vehicles (UAVs) for government and military purposes continues to increase [1], [2]. Small UAVs are often used for delivering information via images or video [3]. Our work's focus is based on algorithms that estimate a three-dimensional (3D) scene or object from two dimensional (2D) images and video. The two major techniques are Structure from Motion (SFM) [4], [5] and Simultaneous Localization and Mapping (SLAM) [6]. UAVs can utilize either of them for a task such as 3D scene understanding.

SLAM and SFM are related: The former is referred as a sequential SFM pipeline [5] in which a 3D structure of a scene is maintained and updated incrementally [7]. It can be viewed as a specific type of SFM [4]. One of the main differences between SLAM and SFM is that the latter uses optical remote sensing imagery to infer the 3D structure of a building and requires cameras [8]. The former needs a LIDAR sensor (and sometimes sonar) to estimate a depth map of the 3D environment [6].

Compared to SFM, SLAM has some drawbacks [6]. First, lasers and sonar are more costly compared to cameras. Second, they are less effective in highly cluttered environments. Third, heavy LIDAR equipment can limit the mobility of airborne systems. These drawbacks make a pipeline of SFM using cameras an economical option for a small, task-oriented UAV [9], [10]. Finally, these are the reasons why our work is based on a non-LIDAR SFM pipeline.

The SFM pipeline usually consists of many steps, several of which can use RANSAC [11]. Such examples are planar homography [12], projective mapping [13], or geometric correction [5]. In planar homography, RANSAC establishes a relation between *keypoints* [8] in pairs of images. It initializes a set containing a minimal number of points for generating a satisfactory model based on some criteria [11]. Then, the algorithm classifies datapoints into a set of relations where the data is consistent with the model [14]. *Inlier* refers to a datum that is consistent with the relation whereas *outlier* is not [14]. In other words, datapoints that correctly match with correspondences are inliers, and those that do not are outliers. If a set of the data consists of all inliers or has a higher ratio of inliers than a certain threshold, RANSAC can fit a model to the set [11].

There have been efforts to improve effectiveness of RANSAC implementations. Due to its brute-force nature, computational cost is one of its major disadvantages, especially when there is a low inlier ratio in the data space [15]. This leads to some fundamental problems of the RANSAC algorithm: what ratio of outliers should be set as a maximal ratio a model can contain [4] or if there is a more efficient way to find the best model when the data's outlier ratio is unknown [11].

We are motivated to explore various implementations of RANSAC variant algorithms as they apply to small, task-oriented UAVs. These UAVs are often consumer-grade and lack significant computing abilities. While the field is shifting to neural networks, small UAVs lack GPU-compute ability. Few studies have focused on suitability of RANSAC given this limitation [8], [16], [17], [18]. This is the principal motivation for our literature survey in RANSAC and its variants. Within the literature, accuracy is the most consistent concern. To improve it, works often try to reduce the stochastic nature of the original RANSAC. The following section will briefly describe the incremental pipeline of SFM and the RANSAC algorithm. Then, we will discuss methodologies and experimental results of several RANSAC variants.

2. Background

The incremental SFM consists of two phases: correspondence search and incremental reconstruction [8]. The *correspondence search* has three sequential steps [18]. It processes a set of images with the goal of finding feature matches between pairs of images [18]. First, *feature extraction* identifies points of interest in the images [8], [17]. Another algorithm determines if the images contain similar, matching features. This step is called *feature matching* [8]. Next, *geometric verification* solves a relation for mapping one image's features onto the other using *projective geometry* [8]. The correspondence will not always be one to one. For example, an image will contain keypoints that are not contained in another [17]. Or, the mappings of the keypoints are incorrect [8]. Thus, matches will have some errors. If mistakes are viewed as outliers, it is possible to apply a RANSAC algorithm to improve the solution, often called a *scene graph* [8], [17].

This process can be used to correct the geometric relationship in *camera homography* [12]. It refers to the process of transferring points from one image to another, and it can infer the transformation between planes of two images [12]. A RANSAC algorithm can also be used to discard the mismatches when fitting a model for *fundamental matrix* estimation [15], [19], [20]. Another example of applying the algorithm is *epipolar geometry* [14], [21]. It determines the projective geometry between two images [12] by finding the relation between them for a moving camera through the *essential matrix* or the fundamental matrix [17].

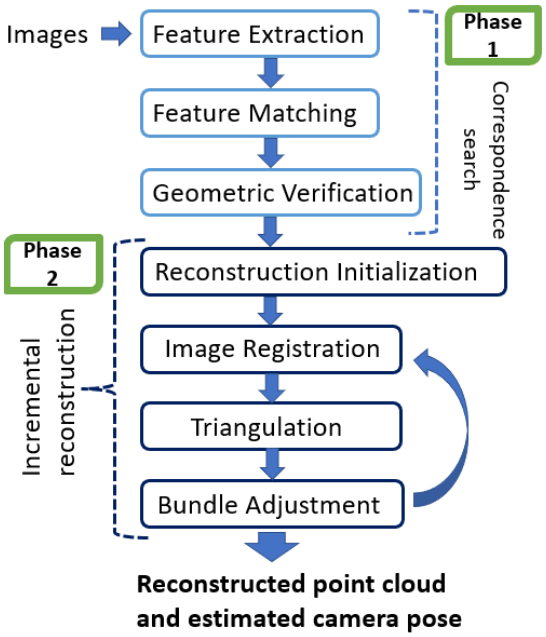

Figure 1. The incremental SFM pipeline [8]

The *incremental reconstruction* phase estimates the camera pose from each image then constructs a point cloud of keypoints in a single 3D coordinate space [8]. As a starting point, the image pair with the most geometrically verified matching points initializes the 3D model (*reconstruction initialization*) [8]. The incremental reconstruction is an iterative process that adds a new image to the 3D model [17]. The pose, position and rotation, relative to the other images, is calculated with a process called *image registration* [8]. New points may then be added to the current model if the triangulated 3D coordinates are observed by, at least, one previously registered image [8], [17]. Finally, *bundle adjustment* minimizes errors across all points [8], [17]. For more detail on the incremental SFM, refer to [8] and [17].

2.1. RANSAC

RANSAC was introduced by Fischler and Bolles in 1981 as a method for automated image analysis [11]. The algorithm seeks candidate solutions by using a small initial number of observations [22]. For example, it discards the mismatches by finding the transformation matrix of feature points [22], [23].

Table 1. The RANSAC algorithm [11], [22]

	The original RANSAC algorithm
	Let D be the set of all datapoints where $\mid D \mid = P$.
	Let S_i be a set of datapoints where $S_i \subset D$, $\mid S_i \mid = N$, and $P \geq N > 0$.
	Let t be the maximum number of iterations.
	Let i be the number of the current iteration.
	Let max be the number of inliers.
	M_i is a model (hypothesis).
1:	$i, max \leftarrow 0$
2:	**while** $(i < t)$ **do**
	// *Hypothesis generation*
3:	$S_i \leftarrow$ Randomly select minimal subset of D
4:	$U_i \leftarrow$ Potential consensus set of S_i
5:	$M_i \leftarrow$ Generate a model from S_i
	// *Hypothesis evaluation*
6:	Calculate error from estimated model
7:	$I_i \leftarrow$ The number of detected inliers
8:	**if** $I_i > max$ **then**
9:	$max \leftarrow I_i$
10:	$U_i^* \leftarrow$ Redefined consensus set
11:	$M_i^* \leftarrow$ Fitted model to U_i^*
12:	**end if**
13:	$i \leftarrow i + 1$
14:	**end while**

In this hypothesize-and-verify framework [14], [22], the size of the subset consists of the minimal number of samples from which a model is hypothesized. The model is then verified against all datapoints in the entire set [14].

Since the inlier ratio is unknown, the lower bound on this ratio is found by using the sample that has the largest number of datapoints consistent with the best quality hypothesis (the *consensus set*) [11]. RANSAC has only one point of termination. Its iterative procedure will continue until the termination criterion is met [11]. In the next section, we examine a partial set of many RANSAC variant algorithms suggested to improve the efficiency of the original RANSAC procedure.

3. Literature Review

Locally optimized RANSAC (LO-RANSAC) modifies RANSAC by applying local optimization to the estimated solution from the random sample [24]. One of the optimization methods presented in this algorithm is the *inner* RANSAC. This method works on a set of inliers where newly sampled datapoints are consistent with the hypothesized model of RANSAC [24]. It applies the RANSAC approach to its *inner* part of the algorithm (see Table 1) once the inliers are selected from the set [24]. It then verifies the model with the entire dataset [24]. Thus, hypotheses do not have to be generated from minimal subsets unlike the original RANSAC [24]. LO-RANSAC increases inliers this way, and consequently decreases the number of drawn samples [24]. In Table 2, we compare LO-RANSAC's and RANSAC's average inlier ratios for two image pairs tested in homography [24]. LO-RANSAC has an average inlier ratio of 0.175, while RANSAC has an average inlier ratio of 0.146 [24].

In Progressive Sample Consensus (PROSAC), a pre-sampling technique is used to obtain inliers in the shortest time by sorting the data in terms of quality [21]. It orders datapoints by quality scores with non-uniform sampling [21]. A quality function is used to weigh the selection procedure. In other words, PROSAC draws samples like RANSAC, but in an order that prioritizes taking the most promising hypotheses [21]. Datapoints are sorted in an order that puts inliers at the top of the list, and the algorithm takes a subset of the data to work with first [21]. PROSAC is tested against RANSAC in their epipolar geometry experiment with three image pairs and the number of datapoints ranged from 250 to 783 [21]. In their experiment, PROSAC runs faster than RANSAC by 41% [21].

Adaptive Real-time Random Sample Consensus (ARRSAC) optimizes model verification and implements improvements to hypothesis generation within the original RANSAC procedure [14]. It uses a partial depth-first pre-emptive strategy to discard bad hypotheses, which reduces the amount of time spent in local optimization [14]. By sampling high-quality models first, ARRSAC generates an adaptively determined number of high-quality hypotheses [14]. This way, ARRSAC improves real-time performance [14]. It also increases the probability of getting a satisfactory solution when the inlier ratios are low [14]. The algorithm is tested on ten image pairs and against six other algorithms [14].

Spatially Consistent Random Sample Consensus (SCRAMSAC) uses a spatial consistency check (SCC) in both model generation and verification [25]. SCRAMSAC pursues a reduced set in a circular region around features in a correspondence pair [25]. The reduced correspondence set has a high inlier ratio [25]. This makes SCRAMSAC more robust than PROSAC and as fast as the latter, which was considered to be the fastest RANSAC variant algorithm at the time [25]. Using SCC, SCRAMSAC does not have to verify model hypotheses against all correspondences while PROSAC does [25]. SCRAMSAC is compared to RANSAC over six image pairs in homography [25].

Because RANSAC becomes slow when estimating models using feature correspondences between images with a low inlier ratio (e.g., <10%), EVSAC is suggested to reduce convergence times [15]. A nearest-neighbor matcher based on *extreme value theory* is used to calculate the confidence that confirms that a datapoint is an inlier [15]. Once the dataset has been evaluated, points with the greatest confidence are sampled first to generate a hypothesis, much like in [21] and [25]. This improves the probability of choosing an all-inlier sample earlier [19]. EVSAC is tested against five RANSAC variant algorithms with nine image pairs for homography estimation and four image pairs for fundamental matrix estimation [15]. In comparison to RANSAC, EVSAC shows an increase of inliers found and reduction in the time taken to find them [15].

Universal RANSAC (USAC) attempts to address many limitations that the original RANSAC has in terms of its accuracy, efficiency, and stability [20]. USAC implements several previously developed strategies: prefiltering, reducing the number of sample models to generate, implementing early termination, and filtering noisy models' post-generation [20]. It is tested against four RANSAC variants in homography, fundamental matrix estimation, and essential matrix estimation [20]. Over five image pairs, USAC shows an average inlier ratio of 0.256 while RANSAC shows an average of 0.229 in homography [20].

Recursive RANSAC (RRANSAC) recursively estimates the parameters of multiple underlying models simultaneously [26]. It can be carried on Unmanned Aerial Systems using a monocular camera for the task of autonomous target following and tracking movements [10]. It is also implemented to UAV Sense and Avoid with a mounted camera [10]. Density-Based Recursive RANSAC (DBR-RANSAC) [27] is a variant algorithm of RRANSAC [26]. DBR-RANSAC increases the accuracy of feature matching of RRANSAC [27]. For each feature, measurement clusters are formed between subsequent images [27]. Then, movement densities for these features are localized [27]. DBR-RANSAC is tested against

Table 2. Each proposed RANSAC variant algorithm's average inlier ratio is compared to the original RANSAC's reported via each corresponding experiment in homography. Next to the number of image pairs, the average number of datapoints taken by the feature matching algorithm used in each experiment is indicated per the RANSAC variant algorithm.

Name of RANSAC variant algorithm	# of image pairs	Average # of datapoints	Proposed RANSAC variant algorithm's average inlier ratio	RANSAC's average inlier ratio	Name of algorithm used for obtaining number of matches
LO-RANSAC [24]	2	127	0.175 ± 0.000	0.146 ± 0.000	7-point or 8-point algorithms
SCRAMSAC [25]	6	369	0.605 ± 0.067	0.362 ± 0.046	Spatial Consistency Check
EVSAC [15]	4	945.75	0.069 ± 0.003	0.047 ± 0.001	SIFT
USAC [20]	5	1172	0.256 ± 0.042	0.229 ± 0.030	PROSAC-based sampling

RRANSAC [27]. In the simulation, it does not produce false associations or traces unlike RRANSAC [27]. At the highest density of clutter, it performs six times faster than RRANSAC [27].

Prior Sampling & Sample Check RANSAC (PSSC-RANSAC) addresses datasets with large outlier ratios and a concern for the computational speed on airborne platforms [28]. It incorporates prior sampling that evaluates texture magnitude, spatial consistency, and feature similarity to determine the quality of feature points. Images are divided into blocks, and each block is evaluated for texture magnitude which is comprised of an image's coarseness, roughness, and contrast [28]. The dataset is then sampled using points with higher quality ranking first [28]. The proposed algorithm is tested against three RANSAC variants over seventy-two images and five synthetic datasets [28]. It is claimed that, through this prior sampling and sample check, quicker convergence, increased robustness, and lower error percentages for mismatches are achieved [28].

Neural-Guided RANSAC (NG-RANSAC) aims to improve model hypothesis search and increase the chance of finding outlier-free minimal sets [29]. Prior information of the number of the best matches is used to train NG-RANSAC [29]. Combined with a neural network that predicts a weight for each datapoint and helps the weights guide the sampling of minimal sets in a self-supervised manner, it outperforms the original RANSAC in fundamental matrix estimation using two image pairs [29].

4. Discussion

Table 2 shows the improved average inlier ratios compared to RANSAC's average in homography. Each is calculated from the experimental results in [15], [20], [24], [25]. We find that homography is the most common step the experiments are conducted in. The main concern of our work is computational complexity. Yet, we could not make a meaningful comparison because of variations in experimental setup. First, different images are used in the experiments. Second, a different number of image pairs is used. Third, a different algorithm is applied to obtain the number of matching points in the image pairs. These challenges make comparing results difficult. Despite these challenges, we find that accuracy has been the chief concern for RANSAC variant algorithms.

From the surveyed RANSAC variants, we also report the following strategies: Local optimization [14], [20], [24], Early termination [20], [24], [28], Partial depth-first evaluation [14], Unfixed number of hypotheses [14], [20], [21], Spatial coherence [25], [27], Probabilistic parametric modeling [15], Ranking-based sampling [15], [20], [21], [27], [28], and Guided sampling [29]. Overall, these strategies improve accuracy, and reduce the computation time when the inlier ratios are low. We believe that higher inlier ratios contribute to better accuracy, and computational savings can be achieved from the non-uniform sampling strategies [20].

5. Conclusion

After reviewing a subset of RANSAC variant algorithms in literature, we find that accuracy has been the most desired aspect of improvement. Non-uniform sampling is the most popular strategy to increase accuracy. This remains the same in both [27] and [28]. However, [28] addresses that images from UAVs often contain high outlier ratios. This assumption suggests a focus on robustness to be considered. Finally, [27] prioritizes UAVs' computational limitations. Therefore, a RANSAC variant algorithm should consider accuracy, performance, and robustness. Given these considerations and the state-of-art RANSAC algorithms, we recommend PSSC-RANSAC [28] for small UAVs. In future work, mission characteristics and the environment should also be considered.

6. References

[1] United States. Naval Aviation Enterprise (NAE), "Navy Aviation Vision 2030-2035," October 2021. [Online]. Available: https://media.defense.gov/2021/Oct/27/2002881262/-1/-1/0/NAVY%AVIATION%20VISION%202030-2035_ FNL.PDF. [Accessed 17 January 2023].

[2] United States. Army Unmanned Aircraft Systems (UAS) Center of Excellence (CoE) (ATZQ-CDI-C) Staff, 2010.

[Online]. Available: https://irp.fas.org/program/collect/ /uas-army.pdf. [Accessed 17 Jan. 2023].

[3] F. Ahmed, J. C. Mohanta, A. Keshari and P. S. Yadav, "Recent Advances in Unmanned Aerial Vehicles: A Review," *Arabian J. for Sci. and Eng.,* vol. 47, no. 7, pp. 7963-7984, 2022.

[4] O. Onur, V. Voroninski, R. Basri and A. Singer, "A Survey on Structure from Motion," *Acta Numerica,* vol. 26, pp. 305-364, 2017.

[5] Y. -M. Wei, L. Kang, B. Yang and L. -D. Wu, "Applications of structure from motion: a survey," *J. Zhejiang Univ. – Sci. C,* vol. 14, pp. 486-494, 2013.

[6] J. Fuentes-Pacheco, J. Ruiz-Ascencio and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artif. Intell. Review,* vol. 43, pp. 51-81, 2015.

[7] A. Locher, M. Havlena and L. V. Gool, "Progressive structure from motion," in *Proc. of the Eur. Conf. on Computer Vision (ECCV)*, pp. 20-35, 2018.

[8] S. Bianco, G. Ciocca and D. Marelli, "Evaluating the Performance of Structure from Motion Pipelines," *J. of Imaging,* vol. 8, 2018.

[9] R. L. Wood et al., "2015 Gorkha Post-Earthquake Reconnaissance of a Historic Village with Micro Unmanned Aerial Systems," *16th World Conf. on Earthquake (16WCEE),* 2017.

[10] J. H. Lee, J. D. Millard, P. C. Lusk and R. W. Beard, "Autonomous target following with monocular camera on UAS using Recursive-RANSAC tracker," *2018 Int. Conf. on Unmanned Aircraft Systems (ICUAS),* pp. 1070-1074, 2018.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. of the ACM,* vol. 24, no. 6, pp. 381-395, 1981.

[12] A. Agarwal, C. V. Jawahar and P. J. Narayanan, "A Survey of Planar Homography Estimation Techniques," 2005. [Online]. Available: https://static.aminer.org/pdf/PDF/000/322/809/planar_homography_accuracy_analysis_and_applications.pdf. [Accessed 17 Jan. 2023].

[13] S. Seitz et al., [Online]. Available: https://www.cs.columbia.edu/~allen/F17/NOTES/homography_pka.pdf.

[14] R. Raguram, J -M. Frahm and M. Pollefeys, "A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus," 2008. [Online]. Available: https://doi.org/10.1007/978-3-540-88688-4_37.

[15] V. Fragoso, P. Sen, S. Rodriguez and M. Turk, "EVSAC: Accelerating Hypotheses Generation by Modeling Matching Scores with Extreme Value Theory," *2013 IEEE Int. Conf. on Computer Vision,* pp. 2472-2479, 2013.

[16] S. Choi, T. Kim, W. Yu, "Performance evaluation of RANSAC family," 1997. [Online]. Available: www.bmva.org/bmvc/2009/Papers/Paper355/Paper355.pdf. [Accessed 17 Jan. 2023].

[17] J. L. Schonberger and J. M. Frahm, "Structure-from-motion revisited," in *IEEE Conf. on Computer vision and Pattern Recognition*, 2016.

[18] S. Jiang, C. Jiang and W. Jiang, "Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools," *ISPRS J. of Photogrammetry and Remote Sensing,* vol. 167, pp. 130-251, 2020.

[19] X. Xiao, Z. Lu and J. -H. Xue, "CLUSAC: Clustering Sample Consensus for Fundamental Matrix Estimation," *2021 IEEE Int. Conf. on Image Processing (ICIP),* pp. 3283-3287, 2021.

[20] R. Raguram, O. Chum, M. Pollefeys, J. Matas and J. -M. Frahm, "USAC: A Universal Framework for Random Sample Consensus," *IEEE Trans. on Pattern Analysis and Machine Intell.,* vol. 35, no. 8, pp. 2022-2038, 2013.

[21] O. Chum and J. Matas, "Matching with PROSAC-progressive sample consensus," *2005 IEEE Comput. Society Conf. on Computer vision and Pattern Recognition (CVPR'05),* vol. 1, pp. 220-226, 2005.

[22] M. Fotouhi, H. Hekmatian, M. A. Kashani-Nezhad and S. Kasaei, "SC-RANSAC: spatial consistency on RANSAC," 26 July 2018. [Online]. Available: https://doi.org/10.1007/s11042-018-6475-6.

[23] X. Zhang and Z. Xie, "Reconstructing 3D Scenes from UAV Images Using a Structure-from-Motion Pipeline," *2018 26th Int. Conf. on Geoinformatics,* pp. 1-6, 2018.

[24] O. Chum, J. Matas and J. Kittler, "Locally Optimized RANSAC," 2003. [Online]. Available: https://doi.org/10.1007/978-3-540-45243-0_31.

[25] T. Sattler, B. Leibe and L. Kobbelt, "SCRAMSAC: Improving RANSAC's efficiency with a spatial consistency filter," *2009 IEEE 12th Int. Conf. on Computer Vision,* pp. 2090-2097, 2009.

[26] P. C. Niedfeldt and R. W. Beard, "Recursive RANSAC: Multiple Signal Estimation with Outliers," *IFAC Proceedings Volumes,* 2013.

[27] F. Yang, W. Tang and H. Lan, "A density-based recursive RANSAC algorithm for unmanned aerial vehicle multi-target tracking in dense clutter," *2017 13th IEEE Int. Conf. on Control & Automation (ICCA),* pp. 23-27, 2017.

[28] J. Zheng, W. Peng, Y. Wang and B. Zhai, "Accelerated RANSAC for Accurate Image Registration in Aerial Video Surveillance," *IEEE Access,* vol. 9, pp. 36775-36790.

[29] E. Brachmann and C. Rother, "Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses," *2019 IEEE/CVF Int. Conf. on Computer Vision (ICCV),* pp. 4321-4330, 2019.

Address for correspondence:

Yeana Lee Bond.
Dept. of Computer & Electrical Engineering & Computer Science
California State University, Bakersfield
9001 Stockdale Hwy.
Bakersfield, CA.
ybond@csub.edu