

New Conditional Sampling Strategies for Speeded-Up RANSAC

Tom Botterill^{1, 2}

tom.botterill@grcnz.com

Steven Mills²

steven.mills@grcnz.com

Richard Green¹

richard.green@canterbury.ac.nz

¹ Department of Computer Science
University of Canterbury
Christchurch 8140, NZ

² Geospatial Research Centre (NZ) Ltd.
Private Bag 4800
Christchurch 8140, NZ

RANSAC (Random Sample Consensus) [2] is a popular algorithm in computer vision for fitting a model to data points contaminated with many gross outliers. Traditionally many small hypothesis sets are chosen randomly; these are used to generate models and the model consistent with most data points is selected. Instead we propose that each hypothesis set chosen is the one most likely to be correct, conditional on the knowledge of those that have failed to lead to a good model. We present two algorithms, BaySAC and SimSAC, to choose this most likely hypothesis set.

A common use for RANSAC is for estimating the essential matrix describing the relative position of two cameras, which can be computed from stereo matches between two images. Given a set of these stereo correspondences the classic RANSAC sampling algorithm selects random subsets of five points, and for each subset computes all possible essential matrices. Each essential matrix is checked against all correspondences until one compatible with a large number of correspondences is found.

RANSAC for essential matrix estimation can be a costly part of real-time Visual Navigation schemes because of the large number of hypothesis sets that must be tried before finding one uncontaminated by outliers. This number may be reduced considerably if outlier probabilities can be estimated, *e.g.* from stereo correspondence match strengths. PROSAC (Progressive Sample Consensus) [1] ranks data points by prior probability then selects subsets in (approximate) order of prior likelihood. Alternatively Guided-MLESAC [3] selects random subsets where each data-point is selected with probability in proportion to its prior inlier likelihood.

These sampling methods fail to take into account is the information gained by testing hypothesis sets and finding them to be contaminated by outliers, unlike the two methods proposed here which are based on the following observation: a hypothesis set leading to a model consistent with few data points probably contains one or more outliers (the alternative possibility is that it contains a degenerate configuration of inliers). Hypothesis sets with one or more data points in common with this set are also now less likely, as they are likely to include the same outlier(s).

Ideally at each time we will choose one of the hypothesis set that is *most likely to contain no outliers* based on the prior probabilities and the history of contaminated samples. This strategy minimises the number of hypotheses that must be tested before finding one consisting entirely of inliers. Unfortunately a closed-form solution for this posterior probability is algebraically intractable. Instead we present two methods of approximating this probability, both of which are shown to work well in practice.

Naïve Bayes method—BaySAC

Our first proposed method, BaySAC, assumes independence between inlier probabilities of data points in the same hypothesis set. First select the data points with the highest inlier probabilities as our hypothesis set. This is the most likely under our assumption. After testing a hypothesis set and finding it was contaminated with outliers update the inlier probabilities of each of these data points by Bayes' rule. Repeat with new probabilities.

This method performs well, outperforming PROSAC and Guided-MLESAC in a range of situations, and has a low computational cost.

Simulation method—SimSAC

An alternative way to compute data point inlier probabilities is by simulation. Repeatedly choose random inlier/outlier statuses for each data point, check whether these statuses are compatible with having failed to observe an inlier set so far, and if so accumulate a histogram of inlier counts for each of the data points. The highest peaks in this histogram form the most likely inlier set. Assuming independence between inlier probabilities for the most likely inliers at time t (a weaker assumption than the Naïve Bayes

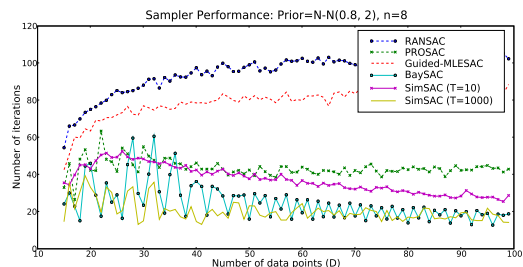


Figure 1: BaySAC and SimSAC reduce the number of iterations needed to find an all-inlier hypothesis set when $N-M$ correspondences are included.



Figure 2: $N-M$ correspondences are useful in some environments

assumptions in BaySAC), and given an arbitrarily large number of samples, this algorithm will give us the actual most likely hypothesis set at each time. SimSAC with a large number (*e.g.* 1000) samples provides a ‘ground truth’ reference—it is unlikely that there is enough information available for any algorithm to perform substantially better than this. While SimSAC is better than other sampling strategies at reducing the number of hypothesis sets needed, in practice it is too slow for most applications. Reducing the number of samples (to 10 for example) gives a strategy with similar performance to BaySAC but which is still more costly.

$N-M$ Stereo Correspondences

Often found that multiple (N) features in one image appear sufficiently similar to multiple (M) features in the other that they could be the same feature. Common approaches to this situation include limiting the image distance between features in the two images, and/or to reject all correspondences with multiple similar-looking potential matches, but in visually poor or self-similar environments too few good correspondences may remain for scene geometry to be recovered accurately. An alternative and effective strategy is to introduce all $N \times M$ possible correspondences (with appropriate inlier probabilities) and to use BaySAC to select those that are inliers. Unlike RANSAC and Guided-MLESAC this does not lead to more hypothesis sets being contaminated by outliers.

Conclusions

BaySAC is an effective way to speed up RANSAC sampling when prior probabilities can be estimated. In some cases the number of iterations required can be halved compared with simple RANSAC sampling, and it significantly outperforms other sampling algorithms that are based on estimated prior probabilities. Performance is good even when prior probability estimates are only approximate.

- [1] O. Chum and J. Matas. Matching with PROSAC - progressive sample consensus. *Proc CVPR*, 2005.
- [2] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM*, 24(6):381–395, 1981.
- [3] Ben J. Tordoff and David W. Murray. Guided-MLESAC: Faster image transformation estimation by using matching priors. *PAMI*, 27(10):1523–1535, 2005.