Efficient Globally Optimal Consensus Maximisation with Tree Search

Tat-Jun Chin*, Pulak Purkait*, Anders Eriksson† and David Suter*
School of Computer Science, The University of Adelaide*, School of Electrical Engineering and Computer Science, Queensland University of Technology†

Maximum consensus is one of the most popular robust criteria for geometric estimation problems in computer vision. Given a set of measurements \( \mathcal{X} = \{ x_i \}_{i=1}^N \), the criterion aims to find the estimate \( \theta \) that agrees with as many of the data as possible (i.e., the inliers) up to a threshold \( \varepsilon \)

\[
\max_{\theta \in \mathcal{X}} \left| \{ i \mid r_i(\theta) \leq \varepsilon \} \right|
\]

subject to \( r_i(\theta) \leq \varepsilon \) \( \forall x_i \in \mathcal{X} \),

where \( r_i(\theta) \) is the residual of \( x_i \). For example, in triangulation we wish to estimate the 3D point \( x \) seen in \( N \) views, where \( \mathcal{X} \) contains the 2D observations \( x_i \) of the point and the associated camera matrices \( P_i \in \mathbb{R}^{3 \times 4} \). The residual \( r_i(\theta) \) is the reprojection error in the \( i \)-th view.

\[
r_i(\theta) = \frac{\| (P_{i,1:2} - x P_{i,3}) \theta \|}{P_{i,3} \theta},
\]

where \( \tilde{\theta} = [\theta^T \ 1]^T \), \( P_{i,1:2} \) is the first-two rows of \( P_i \), and \( P_{i,3} \) is the third row of \( P_i \). The additional constraint \( P_{i,3} \theta > 0 \) must be imposed such that the 3D point lies in front of the cameras. Another example is homography fitting, where a set of point matches \( \mathcal{X} = \{ (u_i^l, u_i^r) \}_{i=1}^N \) across two views, we wish to estimate the homography \( \theta \in \mathbb{R}^{3 \times 3} \) that aligns the points. The residual is

\[
r_i(\theta) = \frac{\| (\theta_{1:2} - u_i^l \theta_3) u_i^r \|}{\theta_3 u_i^r},
\]

where \( \tilde{\theta}_i = [\theta_i^T \ 1]^T \), \( \theta_{1:2} \) is the first-two rows of \( \theta \), and \( \theta_3 \) is the 3rd row of \( \theta \).

In this work, we consider residual functions \( r_i(\theta) \) that are pseudo-convex.

This is known to include many other estimation problems in computer vision. Apart from the above two examples, other problems that involve pseudo-convex residuals include linear regression, camera resectioning, SfM and with known rotations, etc. [4, 7].

Random sample consensus (RANSAC) [2] has been the dominant approach. The method randomly draws \( p \)-tuples from \( \mathcal{X} \), where \( p \) is the minimum number of data to instantiate \( \theta \). The consensus score \( |I| \) of each sample \( \theta \) is obtained, and the \( \theta \) with the highest score and its consensus set \( I \) are returned. A probabilistic bound on the number of samples required can be derived as a stopping criterion.

A major shortcoming of randomised methods such as RANSAC is the lack of absolute certainty that the obtained solution is optimal, or indeed whether it represents a satisfactory approximation at all. A less recognised fact is that even if all \( \binom{P}{p} \) subsets are examined, we may not find the globally optimal solution \( \theta^\star \) to (1), since \( \theta^\star \) does not generally correspond to an estimate from a \( p \)-tuple (see below).

Solving (1) exactly is computationally challenging. Several authors proposed methods based on branch-and-bound (BnB) [5, 10]. However, BnB is typically slow, especially if \( \theta \) is high-dimensional. In fact, RANSAC is suggested to suboptimally preprocess the data (which may cause genuine inliers to be discarded), before BnB is invoked to refine the result. More fundamentally, the BnB methods are specialised for linear residuals. For many vision problems, this entails linearising \( r_i(\theta) \) and adopting algebraic residuals which are not geometrically meaningful.

It has been proven [1, 8] that for various estimation tasks, \( \theta^\star \) can be found as the solution on a subset of \( \mathcal{X} \) of size \( d \), where \( d \geq p \) and \( d \ll N \) (the actual value of \( d \) depends on the particular problem). Both works proposed to find \( \theta^\star \) by exhaustively searching over all \( \binom{P}{d} \) subsets of \( \mathcal{X} \). Although the number of subsets is polynomial w.r.t. \( N \), in realistic problems the number is impractically large. Olsson et al. [8] also proposed using RANSAC with an optimality verification step. However, the fact remains that an enormous number of subsets may need to be sampled.

Due to the much greater computational expense, currently available global methods are not competitive against RANSAC and its variants. In this paper, we make significant progress towards solving (1) exactly and efficiently. Leveraging on the framework of LP-type methods [6, 9], we show how maximising consensus can be cast as a tree search problem. Figure 1(a) illustrates this idea.

We then propose an algorithm based on A* search [3] to traverse the tree. Similar in spirit to [1, 8], we aim to find the optimal data subset. However, instead of sampling or enumerating the subsets, our algorithm deterministically locates the best subset. The A* technique ensures that only a tiny fraction of available subsets need to be explored. Figure 1(b) illustrates our method. Despite its combinatorial nature, our algorithm is fast - on several common estimation problems, our algorithm is orders of magnitude faster than previous exact methods for maximum consensus. Further, our method does not require linearising the residual.

Acknowledgments This work was supported by ARC grants DP130102524 and DE130101775.