REFORMULATING MAXIMUM CONSENSUS AS A SET-HITTING PROBLEM FOR ROBUST ESTIMATION AND OUTLIER REMOVAL UNDER HIGH NOISE

Anonymous authors
Paper under double-blind review

ABSTRACT

Maximum consensus (MAXCON) is a standard criterion for outlier removal and robust model estimation. Computing the *exact* optimal consensus set is a hard optimization problem as the standard big-M mixed-integer linear-programming (MILP) formulation scales poorly with the number of data-points, model dimension, and, most importantly, the percentage of outliers. To address this, we provide an exact reformulation of MAXCON as a *set-hitting* problem, and develop point- and subset-based criteria that remove guaranteed outliers, thereby reducing the size of the reformulated problem. In most cases the reduced problem either admits a tight LP relaxation or its greedy solution is provably optimal, so the exact optimum is certified without solving any MILP – removing the dependence on a licensed solver. In the remaining cases, the original MILP is solved over the much smaller pruned data-set, preserving tractability. On synthetic and real-world datasets with a large number of data-points and 50–80% outliers, the proposed approach recovers the *exact* consensus set while running on average 10–$100\times$ faster than the MILP baseline, which often fails to solve the problem exactly even after exhausting a 48-hour runtime budget.

1 INTRODUCTION

Estimating a reliable model from data contaminated by outliers is a challenge that recurs throughout science and engineering, from robust statistics to computer vision. The maximum-consensus (MAXCON) criterion addresses it by finding the model consistent with the largest subset of the data, and its tolerance to high outlier rates makes it a foundational building block for robust estimation—underpinning various vision tasks ranging from affine estimation to multi-view geometry.

In practice, MAXCON is most often tackled with randomized methods such as RANSAC (Fischler & Bolles, 1981) and its many variants (Torr & Zisserman, 2000; Chum & Matas, 2005; Chum et al., 2003; Raguram et al., 2013; Barath & Matas, 2018; Barath et al., 2020), which are fast but offer no guarantees: they neither certify the consensus set as maximal nor prove any point an outlier. Most importantly, their reliability degrades as the outlier rate and model dimension grow. Interestingly, optimally solving MAXCON under very high outlier rates has largely remained unsolved due to its challenging combinatorial nature. In this paper, we address this by developing a scalable approach to optimally solve MAXCON.

For a dataset $\mathcal{D}$ containing N points, $\mathcal{D} = \{(a_1, b_1), \dots (a_{\mathrm{N}}, b_{\mathrm{N}})\}$ where $a_i \in \mathrm{R}^d$ and $b_i \in \mathrm{R}$, mathematically, following Chin et al. (2016), MAXCON can be written as the following mixed-integer linear-programming (MILP) optimization problem:

$$\mathcal{MC}1 : \min_{\mathcal{Z}, w} \quad \sum_{i=1}^{\mathrm{N}} z_i \tag{1a}$$

$$\text{s.t.} \quad |a_i^T w - b_i| \leq \epsilon + z_i \mathrm{M} \quad \forall \ (a_i, b_i) \in \mathcal{D}, z_i \in \mathcal{Z} \tag{1b}$$

$$\mathcal{Z} \in \{0,1\}^{\mathrm{N}}, w \in \mathrm{R}^d$$

where w denotes the hypothesis (or model parameters), $|a_i^T w - b_i|$ is the resulting residual corresponding to point (a_i, b_i), ϵ denotes the pre-defined acceptable residual tolerance or the inlier

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

threshold and M is a large pre-defined constant. For every (a_i, b_i), the corresponding binary variable z_i decides whether its an inlier (resp. outlier) if $z_i = 0$ (resp. $z_i = 1$). The objective is to find a feasible w such that the number of outliers are minimized.
Solving (1) exactly using an off-the-shelf MILP solver is hard and quickly becomes intractable as the size of the dataset N, problem dimension d and percentage of outliers p_{out} grows. Before developing our approach to exactly solve (1), we first discuss the related work along with their limitations.

Related Work: A large body of work has sought fast (randomized) approximations to (1), mainly different variants of RANSAC as already mentioned previously. A recent line of work learns to reject outlier correspondences directly from data. These include DSAC (Brachmann et al., 2017), NG-RANSAC (Brachmann & Rother, 2019), Ranftl & Koltun (2018), (Zhang et al., 2022) Yi et al. (2018). These methods are appealing due to their differentiable nature, but require in-distribution data and training. In contrast, deterministic (approximate) methods proposed in Le et al. (2021), Purkait et al. (2017) and Cai et al. (2018) start from a randomized point but in each iteration make deterministic updates by solving a relaxed version of the original optimization problem. These methods, although mathematically elegant, require careful selection of parameters governing the degree of relaxation. In summary, all these three related lines of work, i.e. RANSAC based methods, learning based methods and deterministic methods, do not solve (1) exactly or provide a certificate of optimality, and also do not provide any guarantees on the outliers removed.

We now discuss works which aim to solve MAXCON (1) exactly. For a detailed survey, see Chin & Suter (2017). Olsson et al. (2008), Enqvist et al. (2012), Enqvist et al. (2015) and Chin et al. (2018). provide insights and bounds on the NP-hardness of exactly solving MAXCON.
Chin et al. (2015) proposed a tree-search based approach for exactness, but tackles only small instances (N $\leq$ 100) with very low noise ($\leq$ 10%). Cai et al. (2019) proposed a revised version which can handle large N, but still limited to very low noise settings. Speciale et al. (2017) proposed a linear matrix inequality (LMI) based formulation requiring a semi-definite programming(SDP) solver, therefore limited to small instances. A few approaches centre around solving MAXCON as an MILP for global optimality, similar to (1). Li (2009), Zheng et al. (2011) proposed improvements on branch-and-bound (BnB) strategies; Bazin et al. (2013b) adapts BnB over rotation search space and extended to essential matrix by Yang et al. (2014). Bazin et al. (2013a) and Campbell et al. (2020) incorporates feature matching as well with BnB. Chin et al. (2016) proposes GORE, a BnB approach for guaranteed outlier removal by solving a MILP similar to (1) for each data-point (a_i, b_i) by testing its participation against a pre-computed bound in the final inlier-set by forcing $z_i = 0$.

A **major limitation** of all the existing exact approaches is that they do not scale beyond N $\geq$ 100 under high noise setting i.e. $p_{\text{out}} \geq 50\%$. Even with state-of-the-art solvers like Gurobi, the tree induced by BnB search when solving (1) simply becomes intractable. For empirical evidence, see Table 1.

Our novel contributions include:

1. **Exact set-hitting reformulation.** We recast MAXCON as an equivalent *set-hitting* problem and prove the reformulation to be exact via Helly's theorem (Theorem 2.2).
2. **Guaranteed outlier removal.** We derive two provable outlier-detection criteria—a point-consensus-count criterion (Theorem 3.1) and its subset-consensus-count generalization (Theorem 3.2)—that certifiably identify outliers and reduce the size of the reformulated problem.
3. **Inlier upper bound.** We establish a counting result (Corollary 3.3) that upper-bounds the maximum number of inliers, equivalently lower-bounding the number of outliers, directly from consensus counts.
4. **Optimality with orders-of-magnitude speed-ups.** On synthetic and real-world data (affine image matching and epipolar geometry) with N $\geq$ 100 and $p_{\text{out}} = 50$–80%, our approach recovers the *optimal* consensus set while outperforming the Gurobi MILP baseline by one to two orders of magnitude in runtime, also mostly avoiding the out-of-memory failures incurred by the baseline.
5. **Solver-free optimality.** We observe that after outlier removal, the reduced set-hitting instance frequently admits either a tight LP relaxation or a *provably* optimal greedy solution—certifying the global optimum without solving any MILP on 100% of synthetic data instances and around 56% of real-world dataset instances, and thus significantly reducing the dependence on a licensed solver; on the remaining 44%, the original MILP is solved over the much smaller pruned set, preserving tractability.

2 MAXIMUM CONSENSUS AS A SET-HITTING PROBLEM

Let $\mathcal{S}$ denote a non-empty subset of indices: $\mathcal{S} \in \{1, \ldots, \mathrm{N}\}$. Correspondingly $\mathcal{D}_{\mathcal{S}}$ will denote a subset of data-points in $\mathcal{D}$, $\mathcal{D}_{\mathcal{S}} \subseteq \mathcal{D}$, where $\mathcal{D}_{\mathcal{S}} = \{(a_i, b_i) \ \forall \ i \in \mathcal{S}\}$. Similarly, $\mathcal{Z}_{\mathcal{S}}$ will denote a subset of z_i's in $\mathcal{Z}$, $\mathcal{Z}_{\mathcal{S}} \subseteq \mathcal{Z}$, where $\mathcal{Z}_{\mathcal{S}} = \{z_i \ \forall \ i \in \mathcal{S}\}$, corresponding to constraints $|a_i^T w - b_i| \leq \epsilon + z_i \mathrm{M} \ \forall \ (a_i, b_i) \in \mathcal{D}_{\mathcal{S}}$ in (1).
For a given subset $\mathcal{D}_{\mathcal{S}}$, its hypothesis feasibility set $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}})$ can be defined as follows:

$$\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}}) = \left\{ w \in \mathrm{R}^{\mathrm{d}} \ \middle| \ -\epsilon \leq a_i^T w - b_i \leq \epsilon \ \forall \ (a_i, b_i) \ \in \mathcal{D}_{\mathcal{S}} \right\} \tag{2}$$

where emptiness of $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}})$ is sufficient to determine the consensus between points (a_i, b_i) in $\mathcal{D}_{\mathcal{S}}$. For the case when $|\mathcal{D}_{\mathcal{S}}| \leq d$, the system of inequalities: $-\epsilon \leq a_i^T w - b_i \leq \epsilon \ \forall \ (a_i, b_i) \ \in \mathcal{D}_{\mathcal{S}}$ is always feasible for any $\epsilon \geq 0$, simply due to the fact that a hyperplane $w \in \mathrm{R}^d$ (not necessarily unique) passing through any d (or fewer) number of points always exists. Therefore, $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}})$ is always non-empty or equivalently the points in $\mathcal{D}_{\mathcal{S}}$ are always in consensus. However, for the case when $|\mathcal{D}_{\mathcal{S}}| \geq d + 1$, $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}})$ maybe empty ($\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}}) = \phi$). In such cases points in $\mathcal{D}_{\mathcal{S}}$ are not in consensus. Importantly, we claim the following:

Theorem 2.1. *For any subset $\mathcal{D}_{\mathcal{S}} \subseteq \mathcal{D}$, if the corresponding feasibility set $\mathcal{F}(\mathcal{D}_{\mathcal{S}})$ is empty, then the following is a valid constraint (or cut) to MAXCON* (1)*:*

$$\sum_{\forall z_i \in \mathcal{Z}_{\mathcal{S}}} z_i \geq 1 \tag{3}$$

Proof: $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}})$ being empty implies that the points in subset $\mathcal{D}_{\mathcal{S}}$ are not in consensus, therefore these points together cannot be part of any feasible solution to (1) let alone optimal. Importantly, this implies that there is at-least one outlier in $\mathcal{D}_{\mathcal{S}}$. Since, z_i's are binary variables, therefore constraint (3) enforces that at-least one of the z_i's corresponding to a point (a_i, b_i) in $\mathcal{D}_{\mathcal{S}}$ takes a value of 1 or is classified as an outlier in any feasible solution.

To generate a cut of type (3) for a subset $\mathcal{D}_{\mathcal{S}}$, we need to determine if $\mathcal{F}(\mathcal{D}_{\mathcal{S}})$ is empty. This can be easily done using a LP-solver, by evaluating the feasibility of corresponding linear constraints: $-\epsilon \leq a_i^T w - b_i \leq \epsilon \ \forall \ (a_i, b_i) \ \in \mathcal{D}_{\mathcal{S}}$. Note that Theorem 2.1 holds for any arbitrary sized subset $\mathcal{D}_{\mathcal{S}}$, however most of our subsequent discussion will be limited to subsets of size $d + 1$. Therefore, let $\mathcal{S}^{d+1}$ denote a subset of indices in $\mathcal{S}$ of size $d + 1$ and correspondingly $\mathcal{D}_{\mathcal{S}^{d+1}}$ denote a subset of data-points in $\mathcal{D}$ of size $d + 1$. For a dataset $\mathcal{D}$ with N points, $|\mathcal{D}| = \mathrm{N}$, there are $\binom{\mathrm{N}}{d+1}$ different $\mathcal{D}_{\mathcal{S}^{d+1}}$ subsets possible. Let the collection of these be denoted with a universal set $\mathcal{U}^{d+1}$ as following:

$$\mathcal{U}^{d+1} = \left\{ \mathcal{D}_{\mathcal{S}_1^{d+1}}, \mathcal{D}_{\mathcal{S}_2^{d+1}}, \ldots, \mathcal{D}_{\mathcal{S}_i^{d+1}}, \ldots, \mathcal{D}_{\mathcal{S}_{\binom{\mathrm{N}}{d+1}}^{d+1}} \right\} \tag{4}$$

We can further divide the sub-sets in the universal set $\mathcal{U}^{d+1}$ into two complementary (and dis-joint) sets as follows:

$$\mathcal{U}_{\mathcal{I}}^{d+1} = \left\{ \mathcal{D}_{\mathcal{S}_i^{d+1}} \in \mathcal{U}^{d+1} \ \middle| \ \mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}_i^{d+1}}) = \phi \right\}, \qquad \mathcal{U}_{\mathcal{F}}^{d+1} = \mathcal{U}^{d+1} \setminus \mathcal{U}_{\mathcal{I}}^{d+1} \tag{5}$$

where $\mathcal{U}_{\mathcal{I}}^{d+1}$ contain subsets $\mathcal{D}_{\mathcal{S}_i^{d+1}}$ for which the corresponding $d+1$ data-points are **not** in consensus as $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}_i^{d+1}}) = \phi$. Complementarily, $\mathcal{U}_{\mathcal{F}}^{d+1}$ contain subsets where data-points in each subset are in consensus. We can apply Theorem 2.1 on elements from $\mathcal{U}_{\mathcal{I}}^{d+1}$ to generate valid cuts and re-write MAXCON (1) as follows:

$$\begin{aligned}
\mathcal{MC}2 : \min_{z,w} \quad & \sum_{i=1}^{\mathrm{N}} z_i \\
\text{s.t.} \quad & |a_i^T w - b_i| \leq \epsilon + z_i \mathrm{M} \quad \forall \ i \in \{1, \ldots, \mathrm{N}\} && (6) \\
& \sum_{\forall z_i \in \mathcal{Z}_{\mathcal{S}_i^{d+1}}} z_i \geq 1 \quad \forall \quad \mathcal{D}_{\mathcal{S}_i^{d+1}} \in \mathcal{U}_{\mathcal{I}}^{d+1} && (7) \\
& z_i \in \{0, 1\} \quad \forall \ i \in \{1, \ldots, \mathrm{N}\}, w \in \mathrm{R}^d
\end{aligned}$$

Note that since these constraints are valid cuts therefore any integral feasible solution in $\mathcal{MC}1$ is still a feasible solution in $\mathcal{MC}2$.

Theorem 2.2. *Maximum Consensus (MAXCON) can be reformulated as the following set-hitting problem:*

$$\begin{aligned} \mathcal{MC}3: \min_{z} \quad & \sum_{i=1}^{\mathrm{N}} z_i \\ \text{s.t.} \quad & \sum_{\forall z_i \in \mathcal{Z}_{\mathcal{S}_i^{d+1}}} z_i \geq 1 \quad \forall \quad \mathcal{D}_{\mathcal{S}_i^{d+1}} \in \mathcal{U}_{\mathcal{I}}^{d+1} \\ & z_i \in \{0,1\} \quad \forall \ i \in \{1,\ldots,\mathrm{N}\} \end{aligned} \tag{8}$$

The set of constraints in (8) are known as the **set-hitting constraints** and the goal is to find the smallest subset of z_i's which satisfy (or hits) every constraint (or set) arising from elements in the universal set $\mathcal{U}_{\mathcal{I}}^{d+1}$. In a feasible solution, z_i's which take value 1, are collectively referred to as the **hitting-set** (denoted $\mathcal{H}$).

Proof: We start by noting that the main difference between $\mathcal{MC}2$ and $\mathcal{MC}3$ is the absence of big-M constraints (6) in the latter and therefore $\mathcal{MC}3$ can be simply viewed as a relaxation to $\mathcal{MC}2$. Needless to mention, any (integer) feasible solution in $\mathcal{MC}2$ is still a feasible solution in $\mathcal{MC}3$. Feasibility of any solution $\mathcal{Z}$ in $\mathcal{MC}3$ is solely governed by the set-hitting constraints in (8). Therefore, to prove that $\mathcal{MC}3$ is a valid reformulation of MAXCON ($\mathcal{MC}1$), we need to show that for any feasible solution the corresponding data-points classified as inliers are indeed in consensus. For a feasible $\mathcal{Z}$, let $\mathcal{S}_{il}$ and $\mathcal{S}_{ol}$ denote the set of indices of the inliers and outliers respectively, where $\mathcal{S}_{il} = \{i \in \mathcal{S} | z_i = 0\}$ and $\mathcal{S}_{ol} = \mathcal{S} \setminus \mathcal{S}_{il}$. Corresponding set of inlier and outlier data-points can be denoted as $\mathcal{D}_{\mathcal{S}_{il}} = \{(a_i, b_i) \ \forall \ i \ \in \ \mathcal{S}_{il}\}$ and $\mathcal{D}_{\mathcal{S}_{ol}} = \mathcal{D} \setminus \mathcal{D}_{\mathcal{S}_{il}}$.

Mathematically, we need to show $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}_{il}})$ is a non-empty set, i.e. there exists a $w \in \mathrm{R}^d$ for which $|a_i^T w - b_i| \leq \epsilon \ \forall \ (a_i, b_i) \in \mathcal{D}_{\mathcal{S}_{il}}$. Instead of considering a single convex-set formed by all the linear inequalities corresponding to inliers in $\mathcal{D}_{\mathcal{S}_{il}}$ *jointly*, let's focus on the different convex sets formed by each individual inlier data-point *separately*. Using existing notation of $\mathcal{F}_\epsilon(\cdot)$, these convex sets can be easily represented as:

$$\mathcal{F}_\epsilon(\{(a_i,b_i)\}) \ \forall \ (a_i, b_i) \in \mathcal{D}_{\mathcal{S}_{il}}, \tag{9}$$

where $\mathcal{F}_\epsilon(\{(a_i,b_i)\}) = \{w \in \mathrm{R^d} \ \big| \ -\epsilon \leq a_i^T w - b_i \leq \epsilon\}$. Since $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}_{il}}) \neq \phi$ is same as $\bigcap_{i \in \mathcal{S}_{il}} \mathcal{F}_\epsilon(\{(a_i,b_i)\}) \neq \phi$, we choose to prove the latter by mathematically showing that all the individual convex sets $\mathcal{F}_\epsilon(\{(a_i,b_i)\})$ corresponding to inlier data-points (with $z_i = 0$) for a feasible solution $\mathcal{Z}$ have a point in common.

We first show that any $d+1$ sets in (9) have a point in common. Consider any subset of $d+1$ indices of $\mathcal{S}_{il}$ denoted as $\mathcal{S}_{il}^{d+1}$, where $\mathcal{S}_{il}^{d+1} \subseteq \mathcal{S}_{il}$. The corresponding subset of data-points and z_i's can be denoted as $\mathcal{D}_{\mathcal{S}_{il}^{d+1}}$ and $\mathcal{Z}_{\mathcal{S}_{il}^{d+1}}$ respectively, where $\mathcal{D}_{\mathcal{S}_{il}^{d+1}} \subseteq \mathcal{D}_{\mathcal{S}_{il}}$, $\mathcal{Z}_{\mathcal{S}_{il}^{d+1}} \subseteq \mathcal{Z}_{\mathcal{S}_{il}}$ and $\mathcal{Z}_{\mathcal{S}_{il}^{d+1}} = \{0\}^{d+1}$. Since $\mathcal{Z}$ is a feasible solution satisfying (8) therefore $\sum_{\forall z_i \in \mathcal{Z}_{\mathcal{S}_{il}^{d+1}}} z_i \ngeq 1$. This implies that $\mathcal{D}_{\mathcal{S}_{il}^{d+1}} \notin \mathcal{U}_{\mathcal{I}}^{d+1}$, therefore $\mathcal{D}_{\mathcal{S}_{il}^{d+1}} \in \mathcal{U}_{\mathcal{F}}^{d+1}$ as $\mathcal{U}_{\mathcal{F}}^{d+1}$ is the complementary set of $\mathcal{U}_{\mathcal{I}}^{d+1}$. Recall from definition in (5), any subset of $\mathcal{D}$ with $d+1$ data-points which are in consensus belongs to $\mathcal{U}_{\mathcal{F}}^{d+1}$, since $\mathcal{D}_{\mathcal{S}_{il}^{d+1}} \in \mathcal{U}_{\mathcal{F}}^{d+1}$, therefore $d+1$ data-points in $\mathcal{D}_{\mathcal{S}_{il}^{d+1}}$ are also in consensus. Mathematically, $\bigcap_{i \in \mathcal{S}_{il}^{d+1}} \mathcal{F}_\epsilon(\{(a_i,b_i)\}) \neq \phi$, i.e. any $d+1$ sets in (9) have a point in common. We next recall an important result regarding geometry of convex sets known as Helly's Theorem.

Theorem 2.3. *(Helly's Theorem (Helly, 1923)) Consider a finite collection of convex sets: $\mathcal{C}_1, \ldots, \mathcal{C}_n \in R^d$ where $n \geq d+1$. If the intersection of every $d+1$ sets of this collection is non-empty, then all the sets have a non-empty intersection, i.e. $\bigcap_{i=1}^{n} \mathcal{C}_i \neq \phi$.*

We have already proved that every $d+1$ sets in (9) have a point in common or non-empty intersection. From Helly's theorem, it immediately follows that all sets in (9) have a non-empty intersection i.e.:

$$\bigcap_{\forall \ (a_i,b_i) \ \in \ \mathcal{D}_{\mathcal{S}_{il}}} \mathcal{F}_\epsilon(\{(a_i,b_i)\}) \neq \phi \tag{10}$$

Further, since (10) is same as $\mathcal{F}_\epsilon(\mathcal{D}_{\mathcal{S}_{il}}) \neq \phi$, thus completing our proof.

3 OUTLIER REMOVAL

We now develop a (global) outlier removal approach. Since multiple optimal solutions are possible to (1), therefore a point may be an outlier in one optimal solution but an inlier in another. Typically, such points are quite few in number and majority of the outliers are indeed *global* outliers, i.e. an outlier in all optimal solutions. Let n^*_{il} (resp. n^*_{ol}) denote the (optimal) maximum number of inliers (resp. minimum number of outliers), such that $n^*_{il} + n^*_{ol} = \mathrm{N}$. A valid lower-bound (denoted n^{lb}_{il}) to n^*_{il} is assumed to be known and can be very efficiently computed using approximate methods like RANSAC. Proofs of all technical results subsequently discussed in this section, i.e. Theorems 3.1 and 3.2 and Corollary 3.3, are provided in Appendix A.1.1, A.1.2 and A.1.3 respectively.

For a data-point $(a_i, b_i) \in \mathcal{D}$, we define **point-consensus-count**, denoted $\boldsymbol{C}^{d+1}_{\mathcal{F}}((a_i, b_i))$, as the number different subsets of the remainign data-points $\mathcal{D} \setminus \{(a_i, b_i)\}$ of size d which are in consensus with (a_i, b_i). Using $\mathcal{U}^{d+1}_{\mathcal{F}}$, this can be easily computed as follows:

$$\boldsymbol{C}^{d+1}_{\mathcal{F}}((a_i, b_i)) = \sum_{\forall\ \mathcal{D}_{\mathcal{S}^{d+1}_j} \in \mathcal{U}^{d+1}_{\mathcal{F}}} \mathbb{1}_{[\![(a_i, b_i) \in \mathcal{D}_{\mathcal{S}^{d+1}_j}]\!]} \tag{11}$$

where $\mathbb{1}_{[\![\pi]\!]}$ denotes the indicator function which evaluates to 1 if the predicate π holds and 0 otherwise. We now define an outlier removal criterion using $\boldsymbol{C}^{d+1}_{\mathcal{F}}(\cdot)$.

Theorem 3.1. *Let n^{lb}_{il} be a lower-bound to the max. number of inliers n^*_{il}. Then, $(a_i, b_i) \in \mathcal{D}$ is a global outlier if the following holds true:*

$$\boldsymbol{C}^{d+1}_{\mathcal{F}}((a_i, b_i)) < \binom{n^{lb}_{il} - 1}{d} \tag{12}$$

From Theorem 3.1, a data-point (a_i, b_i) is a guaranteed outlier whenever it is in consensus with fewer than $\binom{n^{lb}_{il}-1}{d}$ different d sized subsets of the remaining data-points $\mathcal{D} \setminus (a_i, b_i)$. **Algorithm 1** applies this criterion *repeatedly* to prune $\mathcal{D}$ and $\mathcal{U}^{d+1}_{\mathcal{F}}$ until no outliers are detected.

We extend the idea of point-consensus-count for a single point (a_i, b_i) to a more generalized **subset-consensus-count** (denoted $\boldsymbol{C}^{d+1}_{\mathcal{F}}(\mathcal{D}_{\mathcal{S}^\rho})$), for a ρ sized sub-set of points (denoted $\mathcal{D}_{\mathcal{S}^\rho}$) where $\mathcal{D}_{\mathcal{S}^\rho} \subseteq \mathcal{D}$ and $2 \leq \rho \leq d$, as follows:

$$\boldsymbol{C}^{d+1}_{\mathcal{F}}(\mathcal{D}_{\mathcal{S}^\rho}) = \sum_{\forall\ \mathcal{D}_{\mathcal{S}^{d+1}_j} \in \mathcal{U}^{d+1}_{\mathcal{F}}} \mathbb{1}_{[\![\mathcal{D}_{\mathcal{S}^\rho} \subset \mathcal{D}_{\mathcal{S}^{d+1}_j}]\!]} \tag{13}$$

Theorem 3.2. *Let n^{lb}_{il} be a lower-bound to the max. number of inliers n^*_{il}. Then, there is at-least one global outlier in the subset $\mathcal{D}_{\mathcal{S}^\rho}$ if the following holds true:*

$$\boldsymbol{C}^{d+1}_{\mathcal{F}}(\mathcal{D}_{\mathcal{S}^\rho}) < \binom{n^{lb}_{il} - \rho}{d + 1 - \rho} \tag{14}$$

Algorithm 2 using the criterion in Theorem 3.2, first prunes $\mathcal{U}^{d+1}_{\mathcal{F}}$ *repeatedly* until no outlier containing subsets $\mathcal{D}_{\mathcal{S}^\rho}$ are detected and then finally prunes $\mathcal{D}$.

Before concluding this section, we outline another technical result, i.e. Corollary 3.3 to compute bounds using the point-consensus-counts.

Corollary 3.3. *Let $c_{(1)} \geq c_{(2)} \geq \cdots \geq c_{(|\mathcal{D}|)}$ denote the point-consensus-counts $\boldsymbol{C}^{d+1}_{\mathcal{F}}((a_i, b_i))$ of the data-points in $\mathcal{D}$ arranged in non-increasing order. Then, an upper-bound n^{ub}_{il} (resp. lower-bound n^{lb}_{ol}) to the max. number of inliers n^*_{il} (resp. min. number of outliers n^*_{ol}) can be computed as follows:*

$$\begin{aligned} &n^{ub}_{il} \geq n^*_{il}, \quad \text{where } n^{ub}_{il} \coloneqq \max\Big\{\, m \in \{1, \ldots, |\mathcal{D}|\} \;\Big|\; c_{(m)} \geq \binom{m-1}{d} \,\Big\} \\ &n^{lb}_{ol} \leq n^*_{ol}, \quad \text{where } n^{lb}_{ol} \coloneqq |\mathcal{D}| - n^{ub}_{il} \end{aligned} \tag{15}$$

Algorithm 1: Outlier Removal using point-consensus-count

Input: $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, n_{il}^{lb}
$\mathcal{D}_o \leftarrow \{\}$
repeat
  $n_o \leftarrow |\mathcal{D}_o|$
  for all $(a_i, b_i) \in \mathcal{D}$ **do**
    Compute $\boldsymbol{C}_{\mathcal{F}}^{d+1}((a_i, b_i))$ using $\mathcal{U}_{\mathcal{F}}^{d+1}$
    if $\boldsymbol{C}_{\mathcal{F}}^{d+1}((a_i, b_i)) < \binom{n_{il}^{lb}-1}{d}$ **then**
      $\mathcal{D}_o \leftarrow \mathcal{D}_o \cup (a_i, b_i)$
    end if
  end for
  $\mathcal{D} \leftarrow \mathcal{D} \setminus \mathcal{D}_o$
  $\mathcal{U}_{\mathcal{F}}^{d+1} \leftarrow \{\mathcal{D}_{\mathcal{S}_j^{d+1}} \in \mathcal{U}_{\mathcal{F}}^{d+1} \mid \mathcal{D}_{\mathcal{S}_j^{d+1}} \subseteq \mathcal{D}\}$
until $|\mathcal{D}_o| = n_o$
Output: $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, $\mathcal{D}_o$

Algorithm 2: Outlier Removal using subset-consensus-count

Input: $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, n_{il}^{lb}, ρ
repeat
  $\mathcal{S}_o^{\rho} \leftarrow \{\}$
  for all $\mathcal{D}_{\mathcal{S}^{\rho}} \subseteq \mathcal{D}$ **do**
    Compute $\boldsymbol{C}_{\mathcal{F}}^{d+1}(\mathcal{D}_{\mathcal{S}^{\rho}})$ using $\mathcal{U}_{\mathcal{F}}^{d+1}$
    if $\boldsymbol{C}_{\mathcal{F}}^{d+1}(\mathcal{D}_{\mathcal{S}^{\rho}}) < \binom{n_{il}^{lb}-\rho}{d+1-\rho}$ **then**
      $\mathcal{S}_o^{\rho} \leftarrow \mathcal{S}_o^{\rho} \cup \mathcal{D}_{\mathcal{S}^{\rho}}$
    end if
  end for
  $\mathcal{U}_{\mathcal{F}}^{d+1} \leftarrow \{\mathcal{D}_{\mathcal{S}_j^{d+1}} \in \mathcal{U}_{\mathcal{F}}^{d+1} \mid \mathcal{D}_{\mathcal{S}^{\rho}} \not\subset \mathcal{D}_{\mathcal{S}_j^{d+1}} \ \forall\, \mathcal{D}_{\mathcal{S}^{\rho}} \in \mathcal{S}_o^{\rho}\}$
until $\mathcal{S}_o^{\rho} = \phi$
$\mathcal{D}_o \leftarrow \mathcal{D} \setminus \{\bigcup_{\mathcal{D}_{\mathcal{S}_j^{d+1}} \in \mathcal{U}_{\mathcal{F}}^{d+1}} \mathcal{D}_{\mathcal{S}_j^{d+1}}\}$
$\mathcal{D} \leftarrow \mathcal{D} \setminus \mathcal{D}_o$
Output: $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, $\mathcal{D}_o$

4 Final Algorithm

In this section, we work towards building an algorithm which efficiently solves $\mathcal{MC}3$ using technical results from previous section. First, we recall a well-known greedy heuristic for generating a high-quality feasible solution or a hitting-set (denoted $\mathcal{H}_g$) to the set of constraints in (8).

The *greedy set-hitting* heuristic constructs the hitting-set $\mathcal{H}_g$ one element (or z_i) at a time. At each iteration it selects a z_i that occurs in the largest number of yet un-hit constraints in (8), adds it to $\mathcal{H}_g$, discards every constraint thereby hit, and repeats until no infeasible constraint remains. Being greedy, $\mathcal{H}_g$ is not guaranteed to be a *minimum* hitting set, but in practice it is of high quality and in-fact can be a minimum hitting set if any one of the conditions outlined in Proposition 4.1 are met.

Proposition 4.1 (Optimality Certificates)**.** *Let n_{ol}^{lb} be the lower-bound of Corollary 3.3 and n_{ol}^{lp} the optimal objective of the lp-relaxation of $\mathcal{MC}3$. Then, a hitting set $\mathcal{H}$ of a non-empty $\mathcal{U}_{\mathcal{I}}^{d+1}$ is a* minimum *hitting set, i.e. an optimal solution of $\mathcal{MC}3$, if one of the following holds:*

$$\text{(i)}\ \ |\mathcal{H}| = n_{ol}^{lb}, \qquad \text{(ii)}\ \ |\mathcal{H}| = \lceil n_{ol}^{lp} \rceil, \qquad \text{(iii)}\ \ |\mathcal{H}| = 1 \tag{16}$$

The proof of Proposition 4.1 is provided in Appendix A.1.4.

We now have all the necessary ingredients and put everything together in **Algorithm 3** i.e. the final algorithm. Algorithm 3 first computes consensus subsets $\mathcal{U}_{\mathcal{F}}^{d+1}$ and then proceeds in three stages, highlighted in the corresponding shaded blocks. The first stage (blue) is an iterative prune-and-tighten loop. Starting with $\mathcal{U}_{\mathcal{F}}^{d+1}$ and the initial bound n_{il}^{lb}, it repeatedly invokes Algorithms 1 and 2 to remove guaranteed outliers until no further removal is possible. If the surviving points are already in consensus, i.e. $\mathcal{F}_{\epsilon}(\mathcal{D}) \neq \phi$, the problem is solved and the loop exits. Otherwise the loop attempts to tighten the inlier lower-bound n_{il}^{lb} using approximate methods, i.e. RANSAC and greedy hitting-set heuristic. First RANSAC is run on the (smaller) surviving points $\mathcal{D}$, and if it fails to improve n_{il}^{lb}, then as a precursor to the greedy hitting-set heuristic, the complement of $\mathcal{U}_{\mathcal{F}}^{d+1}$, i.e. $\mathcal{U}_{\mathcal{I}}^{d+1}$, and the outlier lower-bound n_{ol}^{lb} are computed.
Next, a greedy hitting-set $\mathcal{H}_g$ of $\mathcal{U}_{\mathcal{I}}^{d+1}$ is constructed, which is tested for optimality against the available certificates of *(i)* and *(iii)* of Proposition 4.1 – terminating the algorithm immediately if either holds. Otherwise, the inlier lower bound is attempted to be improved with $|\mathcal{D}| - |\mathcal{H}_g|$. An improved inlier bound, either from RANSAC or from $\mathcal{H}_g$, re-enters the pruning step; the loop ends when neither are able to tighten.

The second stage (amber) is entered only if the loop exits without an exact solution. Here, a single LP relaxation of $\mathcal{MC}3$ is solved over the current $\mathcal{U}_{\mathcal{I}}^{d+1}$, and optimality is certified in one of two ways: if the LP solution is integral, it is itself a minimum hitting-set; otherwise, if its rounded-up objective $\lceil n_{ol}^{lp} \rceil$ matches $|\mathcal{H}_g|$, certificate *(ii)* of Proposition 4.1 proves $\mathcal{H}_g$ optimal. In either case the exact solution is obtained – note that so far no MILP has been solved explicitly.

The third stage (green) is the fallback: the exact optimum is achieved by explicitly solving $\mathcal{MC}1$, but over the heavily pruned set $\mathcal{D}$, which is both far smaller than the original data-set and also contains a lower percentage of outliers. We close with a remark on complexity. Note that the fixed cost incurred by Algorithm 3 in computing $\mathcal{U}_{\mathcal{F}}^{d+1}$ is $\mathcal{O}(\mathrm{N}^{d+1})$ feasibility checks, each a small LP of size d, where $d \ll \mathrm{N}$, and trivially parallelisable. For a fixed d this scales polynomially in N, a major gain over the worst-case $\mathcal{O}(2^{\mathrm{N}})$ incurred by BnB search when solving $\mathcal{MC}1$ as a MILP.

Algorithm 3 Final Algorithm

Input: $\mathcal{D}$, n_{il}^{lb}, ρ, ϵ
Compute $\mathcal{U}_{\mathcal{F}}^{d+1}$ and $\mathcal{D}_o \leftarrow \{\}$
loop
  repeat //prune outliers
    $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, $\mathcal{D}'_o \leftarrow$ Algorithm 1 with inputs $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, n_{il}^{lb}
    $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, $\mathcal{D}''_o \leftarrow$ Algorithm 2 with inputs $\mathcal{D}$, $\mathcal{U}_{\mathcal{F}}^{d+1}$, n_{il}^{lb}, ρ
    $\mathcal{D}_o \leftarrow \mathcal{D}_o \cup \mathcal{D}'_o \cup \mathcal{D}''_o$
  until $\mathcal{D}''_o = \phi$
  break if $\mathcal{F}_\epsilon(\mathcal{D}) \neq \phi$.
  $n_{il}^{r} \leftarrow$ compute no. of inliers via ransac on $\mathcal{D}$
  if $n_{il}^{r} > n_{il}^{lb}$ **then**
    $n_{il}^{lb} \leftarrow n_{il}^{r}$ **and continue** //prune again with improved bound
  end if
  Compute $\mathcal{U}_{\mathcal{I}}^{d+1}$ using $\mathcal{U}_{\mathcal{F}}^{d+1}$ and n_{ol}^{lb} using Corollary 3.3
  $\mathcal{H}_g \leftarrow$ greedy hitting-set of $\mathcal{U}_{\mathcal{I}}^{d+1}$
  if $|\mathcal{H}_g| = n_{ol}^{lb}$ **or** $|\mathcal{H}_g| = 1$ **then**
    $\mathcal{D}, \mathcal{D}_o \leftarrow$ Update with $\mathcal{H}_g$; **break**
  end if
  $n_{il}^{g} \leftarrow |\mathcal{D}| - |\mathcal{H}_g|$, compute no. of inliers via greedy hitting-set $\mathcal{H}_g$
  if $n_{il}^{g} > n_{il}^{lb}$ **then**
    $n_{il}^{lb} \leftarrow n_{il}^{g}$ **and continue** //prune again with improved bound
  else
    break //neither bound improved
  end if
end loop
if $\mathcal{F}_\epsilon(\mathcal{D}) = \phi$ and $|\mathcal{U}_{\mathcal{I}}^{d+1}| \leq$ max_lp_size **then**
  $\mathcal{H}_{lp}$, $n_{ol}^{lp} \leftarrow$ Solve lp-relaxation of $\mathcal{MC}3$ on $\mathcal{U}_{\mathcal{I}}^{d+1}$
  if $\mathcal{H}_{lp} \in \{0,1\}^{|\cdot|}$ (i.e. lp-relax. is tight) **or** $|\mathcal{H}_g| = \lceil n_{ol}^{lp} \rceil$ **then**
    $\mathcal{D}, \mathcal{D}_o \leftarrow$ Update with $\mathcal{H}_{lp}$ or $\mathcal{H}_g$ accordingly
  end if
end if
if $\mathcal{F}_\epsilon(\mathcal{D}) = \phi$ **then**
  $\mathcal{D}, \mathcal{D}_o \leftarrow$ Update with solution of $\mathcal{MC}1$ on $\mathcal{D}$
end if
Output: $\mathcal{D}, \mathcal{D}_o$

5 Experiments

We now evaluate our approach, Algorithm 3 (shorthanded as Algo. 3 henceforth), on synthetic and real-world datasets, benchmarking against $\mathcal{MC}1$ solved with Gurobi v13.0.1 (Gurobi Optimization, LLC, 2026). For both methods—Algo. 3 and directly solving $\mathcal{MC}1$—each instance is run on a machine with 32 CPU cores (Intel® Xeon® Platinum 8488C) and 450 GB of memory, with a time limit of 48 h. An instance that exhausts this budget is marked *timeout* (hits the 48 h limit) or *OOM* (exceeds the memory limit). The value of M used when solving $\mathcal{MC}1$ is decided from the maximum RANSAC residual, which stays below 10^3 in all our cases; to keep the formulation tight, we choose $\mathrm{M} = 100$ when the maximum residual is below 100 and $\mathrm{M} = 1000$ otherwise. We use $\rho = d$ when solving with Algo. 3. We use freely available GLPK (Makhorin, 2026) to solve LPs.

Table 1: Synthetic data runtime (in secs.) comparison: Algo. 3 (Ours) vs. $\mathcal{MC}1$ (MILP).

d	N	50% Outliers			60% Outliers			70% Outliers			80% Outliers		
		Algo. 3 (Ours)	$\mathcal{MC}1$ (MILP)	Gain	Algo. 3 (Ours)	$\mathcal{MC}1$ (MILP)	Gain	Algo. 3 (Ours)	$\mathcal{MC}1$ (MILP)	Gain	Algo. 3 (Ours)	$\mathcal{MC}1$ (MILP)	Gain
2	100	1.6	0.2	0.1×	1.4	0.4	0.3×	1.3	0.5	0.4×	1.3	1.8	**1.4**×
	250	4.5	6.8	**1.5**×	4.1	10.3	**2.5**×	3.1	20.0	**6.4**×	3.8	40.6	**10.7**×
	500	23.1	68.1	**2.9**×	20.4	155.8	**7.7**×	18.9	959.8	**50.7**×	19.5	4046.1	**207.2**×
	750	77.4	381.4	**4.9**×	72.2	6219.5	**86.2**×	64.6	2.3×10^4	**355.4**×	59.4	3.1×10^4	**527.5**×
	1000	188.2	4347.5	**23.1**×	171.0	5.7×10^4	**333.7**×	154.1	1.4×10^5	**930.1**×	147.1	6.7×10^4	**457.0**×
3	100	5.7	4.9	0.9×	5.4	14.2	**2.6**×	5.5	58.1	**10.6**×	5.5	150.4	**27.1**×
	200	69.4	199.4	**2.9**×	69.1	1569.3	**22.7**×	64.9	1.7×10^4	**269.5**×	63.5	6.2×10^4	**980.5**×
	300	368.8	2986.7	**8.1**×	354.6	5.9×10^4	**167.1**×	336.2	9.2×10^4	**272.3**×	318.3	6.8×10^4	**213.8**×
	400	1179.3	2.7×10^4	**22.8**×	1152.7	1.7×10^5	**150.2**×	1130.0	1.1×10^5	**99.2**×	1093.5	9.2×10^4	**84.4**×
4	100	101.3	48.2	0.5×	90.2	354.6	**3.9**×	86.5	1129.4	**13.1**×	88.2	1843.1	**20.9**×
	125	277.6	259.1	0.9×	292.1	1911.3	**6.5**×	261.2	6366.6	**24.4**×	271.7	9711.5	**35.7**×
	150	694.9	820.8	**1.2**×	700.4	9097.5	**13.0**×	689.0	3.4×10^4	**49.7**×	675.1	4.9×10^4	**72.7**×
	175	1533.5	2645.3	**1.7**×	1537.0	5.0×10^4	**32.7**×	1547.4	7.8×10^4	**50.1**×	1489.1	4.0×10^4	**27.0**×

Synthetic data. We generate d-dimensional linear-regression instances for $d \in \{2, 3, 4\}$. For each dimension we fix a ground-truth hyperplane w; the inliers have their first $d-1$ coordinates drawn uniformly in a box of half-width 4, with the remaining coordinate determined by w up to a Gaussian perturbation of scale 0.01, while the outliers are drawn uniformly at random in the same box. To ensure a high-noise regime, the percentage of outliers p_{out} ranges from 50% to 80%. For each configuration (i.e. $d, \text{N}, p_{\text{out}}$) we generate with 5 random seeds. An inlier threshold $\epsilon = 0.2$ is used. Figure 1 illustrates a $d = 2$ instance. **Table 1** reports the mean runtimes. These **results** indicate that Algo. 3 is largely agnostic to the outlier fraction, whereas $\mathcal{MC}1$ degrades sharply as either the outlier fraction or the dimension grows, reaching hundreds- to thousand-fold slow-downs. Importantly, for all the instances Algo. 3 achieves global optimality without ever needing to fall back to solving an MILP with a solver (Gurobi).

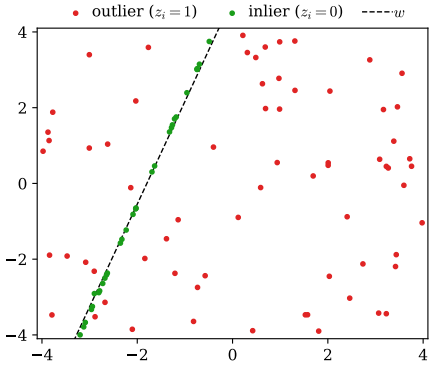


Figure 1: Maximum-consensus line fit ($d = 2$). Inliers ($z_i = 0$, green) lie within $\pm\epsilon$ of the fit w; the rest are outliers ($z_i = 1$, red).

Real data. We now evaluate on real images over two vision tasks, i.e. affine image matching and epipolar (fundamental-matrix) estimation, representative pairs are shown in Figure 2. In both, we detect SIFT features for each image pair (I and I'), form putative correspondences by ratio-test matching and retaining pairs with high noise using an inlier threshold of $\epsilon = 1$ pixel throughout. A correspondence, pairs a point $(x_i, y_i) \in I$ with its match $(x'_i, y'_i) \in I'$. In homogeneous coordinate system: $\tilde{x}_i = [x_i, y_i, 1]^\top$ and $\tilde{x}'_i = [x'_i, y'_i, 1]^\top$, the two tasks are governed, respectively, by a 2D affine map $\tilde{x}'_i = A\,\tilde{x}_i$ and the epipolar constraint $\tilde{x}_i'^\top F\,\tilde{x}_i = 0$, with

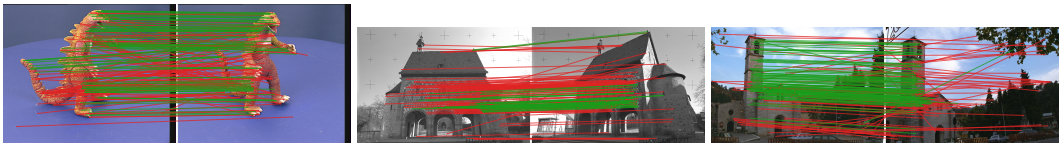
$$A = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (17) \qquad ; \qquad F = \begin{bmatrix} 0 & 0 & f_1 \\ 0 & 0 & f_2 \\ f_3 & f_4 & f_5 \end{bmatrix} \quad (18)$$

each reducing to an instance of (1) as detailed below.

(i) Affine image matching. In this task the two views are related by the affine map $\tilde{x}'_i = A\,\tilde{x}_i$ of (17), with active rows $\beta_1 = [a_{11},\ a_{12},\ t_x]^\top$ and $\beta_2 = [a_{21},\ a_{22},\ t_y]^\top \in \mathrm{R}^3$. From input $\tilde{x}_i$, the two target coordinates (x'_i, y'_i) are linked as: $x'_i = \beta_1^\top \tilde{x}_i$ and $y'_i = \beta_2^\top \tilde{x}_i$; this can therefore be considered as a multi-output instance of linear regression. Following Chin et al. (2016), using ℓ_∞ residual criterion, a correspondence is in consensus when $\max\big(|x'_i - \beta_1^\top \tilde{x}_i|,\ |y'_i - \beta_2^\top \tilde{x}_i|\big) \le \epsilon$, i.e. both coordinates agree to within ϵ. This yields the two constraints: $|x'_i - \beta_1^\top \tilde{x}_i| \le \epsilon + z_i \mathrm{M}$ and $|y'_i - \beta_2^\top \tilde{x}_i| \le \epsilon + z_i \mathrm{M}$, coupled with a single binary variable $z_i \in \{0, 1\}$. Although β_1, β_2 together contribute six parameters, due to coupling the effective dimension of the problem is $d = 3$ for our



(a) boat 2→6 (69% outliers) (b) graf 4→5 (65% outliers) (c) wall 3→5 (78% outliers)



(d) dino 4→7 (60% outliers) (e) kapel 15→25 (67% outliers) (f) valbonne 8→10 (46% outliers)

Figure 2: Representative image pairs from real-data experiments, with inlier correspondences shown in green and outliers in red; (a)–(c) are affine image matching and (d)–(f) affine epipolar geometry.

approach. A detailed reduction is provided in Appendix A.1.5. We evaluate on a total of 40 pairs of images from Oxford's affine-covariant dataset: `boat`, `graf`, and `wall` and the overall runtimes are summarised in **Table 2**. Per-image-pair breakdown is provided in Tables 4–6 in Appendix A.2. These **results** indicate that Algo. 3 scales gracefully, solving 38 of 40 setups—while $\mathcal{MC}1$ exhausts the memory budget on 30 of them and solves just 9; avg. gain varies from 22× to over 200×. Also, Algo. 3 was able to solve 10 of 40 setups without needing a MILP solver.

(ii) Affine epipolar geometry. For this task, correspondences should satisfy the epipolar constraint $\tilde{x}_i'^{\top} F \tilde{x}_i = 0$ with the fundamental matrix F of (18) having 4 degrees of freedom as it is defined only up to scale. Expanding the constraint gives a linear equation, $f_1 x_i' + f_2 y_i' + f_3 x_i + f_4 y_i + f_5 = 0$, an instance of (1) with $d = 4$. We evaluate on frame pairs from the `dino`, `kapel`, and `valbonne` sequences of Oxford's VGG multi-view dataset and report the per-dataset results in **Table 3**. These **results** indicate that Algo. 3 remains reliable at scale, certifying the optimum on all 63 setups—whereas $\mathcal{MC}1$ solves only 29 and fails on the remaining hard instances; avg. gain varies from 23× to 55×. Importantly, Algo. 3 optimally solves 48 of 63 setups without invoking a MILP solver. Per-setup results are given in Tables 7–9 of Appendix A.2.

Table 2: Runtime comparison on affine image matching task: Algo. 3 (Ours) vs. $\mathcal{MC}1$ (MILP).

Dataset	No. of Setups	N range	Outlier range	Algo. 3 avg. (s)	$\mathcal{MC}1$ avg. (s)	Gain avg.	Algo. 3			$\mathcal{MC}1$		
							solved	timeout	OOM	solved	timeout	OOM
boat	6	208–357	52–69	743.0	2.5×10^4	**22.4**×	6	0	0	6	0	0
graf	14	161–396	60–80	13494.9	9.4×10^4	**189.5**×	12	1	1	1	1	12
wall	20	157–380	60–83	1832.2	7.8×10^4	**232.2**×	20	0	0	2	0	18
Total	40	157–396	52–83	5750.8	7.6×10^4	**185.8**×	38/40 (95%)	1/40 (2.5%)	1/40 (2.5%)	9/40 (22.5%)	1/40 (2.5%)	30/40 (75%)

Table 3: Runtime comparison on affine epipolar geometry task: Algo. 3 (Ours) vs. $\mathcal{MC}1$ (MILP).

Dataset	No. of Setups	N range	Outlier range	Algo. 3 avg. (s)	$\mathcal{MC}1$ avg. (s)	Gain avg.	Algo. 3			$\mathcal{MC}1$		
							solved	timeout	OOM	solved	timeout	OOM
dino	26	125–202	43–78	1925.2	6.8×10^4	**54.7**×	26	0	0	9	3	14
kapel	18	126–197	41–67	2253.7	4.1×10^4	**22.6**×	18	0	0	11	1	6
valbonne	19	125–198	42–69	2463.0	6.1×10^4	**32.9**×	19	0	0	9	2	8
Total	63	125–202	41–78	2181.2	5.8×10^4	**38.9**×	63/63 (100%)	0/63 (0%)	0/63 (0%)	29/63 (46%)	6/63 (10%)	28/63 (44%)

6 Conclusion

We reformulated maximum consensus as an exact *set-hitting* problem and derived guaranteed outlier-removal criteria that reduce the original problem substantially—often certifying the optimum *without* any MILP solver. On synthetic and real vision tasks with large number of data-points and up to 80% outliers, our method obtains the exact consensus set while being 10–100× faster than the MILP baseline.

AI USE STATEMENT

In this work, we used generative AI tools to polish the writing of text that we had ourselves drafted, and to assist with LaTeX typesetting of the tables and figures. We have not used generative AI tools for research ideation, for deriving or proving any of the theoretical results, or for designing the algorithm and the experimental protocol; these, and the accompanying implementation, are entirely our own. Nor did we use them for literature search or to produce the bibliography, which was compiled directly from the publishers' official pages for the cited works. We have reviewed all AI-assisted work: every AI-polished passage was checked by the authors against the underlying technical content, and all reported numbers were verified against our own experiment logs. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Each theoretical result is stated with its assumptions made explicit, and complete proofs of Theorems 3.1 and 3.2, Corollary 3.3, and Proposition 4.1 are given in Appendix A.1, with the reduction of affine image matching to our formulation derived in Appendix A.1.5. We will release our code publicly upon acceptance. Our method is specified in full as Algorithm 3, and Section 5 documents the data-generation protocol for the synthetic instances, the feature-matching protocol and inlier threshold for the real image pairs, and the hardware, solver version, time and memory limits, and the choice of M and ρ used for every run. Both real-world benchmarks are publicly available—the affine-covariant regions dataset at `https://www.robots.ox.ac.uk/~vgg/research/affine/` and the multi-view dataset at `https://www.robots.ox.ac.uk/~vgg/data/mview/`—and per-setup results for all 103 setups (40 affine image matching and 63 affine epipolar geometry) are reported in Appendix A.2 rather than aggregate numbers alone.

REFERENCES

Daniel Barath and Jiri Matas. Graph-cut ransac. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6733–6741, 2018. doi: 10.1109/CVPR.2018.00704.

Daniel Barath, Jana Noskova, Maksym Ivashechkin, and Jiří Matas. MAGSAC++, a fast, reliable and accurate robust estimator. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1301–1309, 2020. doi: 10.1109/CVPR42600.2020.00138.

Jean-Charles Bazin, Hongdong Li, In So Kweon, Cédric Demonceaux, Pascal Vasseur, and Katsushi Ikeuchi. A branch-and-bound approach to correspondence and grouping problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1565–1576, 2013a. doi: 10.1109/TPAMI.2012.264.

Jean-Charles Bazin, Yongduek Seo, and Marc Pollefeys. Globally optimal consensus set maximization through rotation search. In Kyoung Mu Lee, Yasuyuki Matsushita, James M. Rehg, and Zhanyi Hu (eds.), *Computer Vision – ACCV 2012*, pp. 539–551, Berlin, Heidelberg, 2013b. Springer Berlin Heidelberg. ISBN 978-3-642-37444-9.

Eric Brachmann and Carsten Rother. Neural-guided ransac: Learning where to sample model hypotheses. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4321–4330, 2019. doi: 10.1109/ICCV.2019.00442.

Eric Brachmann, Alexander Krull, Sebastian Nowozin, Jamie Shotton, Frank Michel, Stefan Gumhold, and Carsten Rother. Dsac — differentiable ransac for camera localization. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2492–2500, 2017. doi: 10.1109/CVPR.2017.267.

Zhipeng Cai, Tat-Jun Chin, Huu Le, and David Suter. Deterministic consensus maximization with biconvex programming. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

Zhipeng Cai, Tat-Jun Chin, and Vladlen Koltun. Consensus maximization tree search revisited. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Dylan Campbell, Lars Petersson, Laurent Kneip, and Hongdong Li. Globally-optimal inlier set maximisation for camera pose and correspondence estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):328–342, 2020. doi: 10.1109/TPAMI.2018.2848650.

Tat-Jun Chin and David Suter. *The Maximum Consensus Problem: Recent Algorithmic Advances*. Synthesis Lectures on Computer Vision. Springer Cham, 2017. ISBN 978-3-031-00690-6. doi: 10.1007/978-3-031-01818-3.

Tat-Jun Chin, Pulak Purkait, Anders Eriksson, and David Suter. Efficient globally optimal consensus maximisation with tree search. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2413–2421, 2015. doi: 10.1109/CVPR.2015.7298855.

Tat-Jun Chin, Yang Heng Kee, Anders Eriksson, and Frank Neumann. Guaranteed outlier removal with mixed integer linear programs. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5858–5866, 2016. doi: 10.1109/CVPR.2016.631.

Tat-Jun Chin, Zhipeng Cai, and Frank Neumann. Robust fitting in computer vision: easy or hard? In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

O. Chum and J. Matas. Matching with prosac - progressive sample consensus. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pp. 220–226 vol. 1, 2005. doi: 10.1109/CVPR.2005.221.

Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In Bernd Michaelis and Gerald Krell (eds.), *Pattern Recognition*, pp. 236–243, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. ISBN 978-3-540-45243-0.

Olof Enqvist, Erik Ask, Fredrik Kahl, and Kalle Åström. Robust fitting for multiple view geometry. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid (eds.), *Computer Vision – ECCV 2012*, pp. 738–751, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-33718-5.

Olof Enqvist, Erik Ask, Fredrik Kahl, and Kalle Åström. Tractable algorithms for robust model estimation. *International Journal of Computer Vision*, 112(1):115–129, 2015. doi: 10.1007/s11263-014-0760-2.

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, June 1981. ISSN 0001-0782. doi: 10.1145/358669.358692. URL https://doi.org/10.1145/358669.358692.

Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2026. URL https://www.gurobi.com.

Eduard Helly. Über mengen konvexer körper mit gemeinschaftlichen punkten. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 32:175–176, 1923.

Huu Le, Tat-Jun Chin, Anders Eriksson, Thanh-Toan Do, and David Suter. Deterministic approximate methods for maximum consensus robust fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):842–857, 2021. doi: 10.1109/TPAMI.2019.2939307.

Hongdong Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *2009 IEEE 12th International Conference on Computer Vision*, pp. 1074–1080, 2009. doi: 10.1109/ICCV.2009.5459398.

Andrew Makhorin. GLPK (GNU Linear Programming Kit), 2026. URL https://www.gnu.org/software/glpk/.

Carl Olsson, Olof Enqvist, and Fredrik Kahl. A polynomial-time bound for matching and registration with outliers. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008. doi: 10.1109/CVPR.2008.4587757.

Pulak Purkait, Christopher Zach, and Anders P. Eriksson. Maximum consensus parameter estimation by reweighted ℓ_1 methods. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2017. URL https://api.semanticscholar.org/CorpusID:4772190.

Rahul Raguram, Ondrej Chum, Marc Pollefeys, Jiri Matas, and Jan-Michael Frahm. Usac: A universal framework for random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):2022–2038, 2013. doi: 10.1109/TPAMI.2012.257.

Rene Ranftl and Vladlen Koltun. Deep fundamental matrix estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

Pablo Speciale, Danda P. Paudel, Martin R. Oswald, Till Kroeger, Luc Van Gool, and Marc Pollefeys. Consensus maximization with linear matrix inequality constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4940–4948, 2017. doi: 10.1109/CVPR.2017.525.

Philip H.S. Torr and Andrew Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000. doi: 10.1006/cviu.1999.0832.

Jiaolong Yang, Hongdong Li, and Yunde Jia. Optimal essential matrix estimation via inlier-set maximization. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *Computer Vision – ECCV 2014*, pp. 111–126, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10590-1.

Kwang Moo Yi, Eduard Trulls, Yuki Ono, Vincent Lepetit, Mathieu Salzmann, and Pascal Fua. Learning to find good correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Jiahui Zhang, Dawei Sun, Zixin Luo, Anbang Yao, Hongkai Chen, Lei Zhou, Tianwei Shen, Yurong Chen, Long Quan, and Hongen Liao. OANet: Learning Two-View Correspondences and Geometry Using Order-Aware Network . *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 44(06):3110–3122, June 2022. ISSN 1939-3539. doi: 10.1109/TPAMI.2020.3048013. URL https://doi.ieeecomputersociety.org/10.1109/TPAMI.2020.3048013.

Yinqiang Zheng, Shigeki Sugimoto, and Masatoshi Okutomi. Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint. In *CVPR 2011*, pp. 1825–1832, 2011. doi: 10.1109/CVPR.2011.5995640.

A Appendix

A.1 Proofs

A.1.1 Proof of Theorem 3.1

Proof: Since n_{il}^{lb} is lower-bound, therefore there exist a set of at-least n_{il}^{lb} number of points in consensus. A point (a_j, b_j) in this set is in consensus with the remaining $n_{il}^{lb} - 1$ number of points. From these $n_{il}^{lb} - 1$ remaining points, we can choose $\binom{n_{il}^{lb}-1}{d}$ number of different d sized subsets, where each of these subsets are in consensus with (a_j, b_j) resulting in subsets of size $d+1$. Therefore, any inlier in the optimal solution has to be in consensus with at-least $\binom{n_{il}^{lb}-1}{d}$ number of different d sized subsets. Mathematically, this can be written as an inequality as done in (12).

A.1.2 Proof of Theorem 3.2

Proof: The argument generalizes that of Theorem 3.1 (the $\rho = 1$ case). Suppose all ρ points of $\mathcal{D}_{\mathcal{S}^\rho}$ are inliers of an optimal consensus set, which contains at-least n_{il}^{lb} points. Then each of the remaining $n_{il}^{lb} - \rho$ inliers is in consensus with $\mathcal{D}_{\mathcal{S}^\rho}$, and any $d + 1 - \rho$ of them together with $\mathcal{D}_{\mathcal{S}^\rho}$ form a $(d+1)$-sized subset that is in consensus. There are $\binom{n_{il}^{lb}-\rho}{d+1-\rho}$ such subsets, implying

$\boldsymbol{C}_{\mathcal{F}}^{d+1}(\mathcal{D}_{\mathcal{S}^\rho}) \geq \binom{n_{il}^{lb}-\rho}{d+1-\rho}$. Contrapositively, if $\boldsymbol{C}_{\mathcal{F}}^{d+1}(\mathcal{D}_{\mathcal{S}^\rho}) < \binom{n_{il}^{lb}-\rho}{d+1-\rho}$, not all ρ points can be inliers simultaneously, i.e. $\mathcal{D}_{\mathcal{S}^\rho}$ contains at-least one global outlier, as done in (14).

A.1.3 Proof of Corollary 3.3

Proof: Let us recall from Theorem 3.1 that any inlier (a_i, b_i) of an optimal consensus set satisfies $\boldsymbol{C}_{\mathcal{F}}^{d+1}((a_i, b_i)) \geq \binom{n_{il}^{lb}-1}{d}$, an implication the theorem applies contrapositively to certify individual data-points as outliers. Corollary 3.3 runs the same implication in the opposite direction: instead of using a lower-bound on the consensus size to prune points, it uses the counts of the points to bound that size itself.

The predicate $c_{(m)} \geq \binom{m-1}{d}$ in (15) can be read as a feasibility test, asking whether a consensus set of size m could exist at all: were there one, each of its m inliers would be in consensus with at-least $\binom{m-1}{d}$ different d sized subsets of the remaining points, forcing the m-th largest count to meet that threshold. Wherever the predicate fails, no consensus set of size m is possible, and n_{il}^{ub} is the largest size that is not ruled out in this way. Its two sides move in opposite directions — $c_{(m)}$ is non-increasing in m by construction, whereas $\binom{m-1}{d}$ is non-decreasing in m, and strictly increasing once $m \geq d+1$ — so it holds up to a single crossing point and fails beyond it, and n_{il}^{ub} is precisely that crossing point.

Formally, the implication holds for any valid lower-bound to n_{il}^*, in particular for n_{il}^* itself. Each of the n_{il}^* inliers of an optimal consensus set therefore has a point-consensus-count of at-least $\binom{n_{il}^*-1}{d}$, i.e. at-least n_{il}^* data-points of $\mathcal{D}$ attain this count. Since the $c_{(m)}$ are arranged in non-increasing order, the statement "at-least m data-points have a count of at-least T" is equivalent to $c_{(m)} \geq T$; taking $m = n_{il}^*$ and $T = \binom{n_{il}^*-1}{d}$ gives $c_{(n_{il}^*)} \geq \binom{n_{il}^*-1}{d}$. Hence n_{il}^* belongs to the set in (15) — which is therefore non-empty and its maximum well-defined — and so $n_{il}^{ub} \geq n_{il}^*$. Consequently the outlier count obeys $n_{ol}^* = |\mathcal{D}| - n_{il}^* \geq |\mathcal{D}| - n_{il}^{ub} = n_{ol}^{lb}$, i.e. n_{ol}^{lb} is a valid lower-bound on n_{ol}^*. Since the point-consensus-counts are already computed when applying Theorem 3.1, obtaining n_{il}^{ub} costs no more than sorting them once and scanning over m.

A.1.4 Proof of Proposition 4.1

Proof: Note that $n_{ol}^{lb}, \lceil n_{ol}^{lp} \rceil$ and 1 are all valid integer lower-bounds to $\mathcal{MC}3$ of varying tightness. Therefore a hitting-set $\mathcal{H}$ of $\mathcal{U}_{\mathcal{I}}^{d+1}$ whose cardinality coincides with any of the three lower-bounds is guaranteed optimal since $\mathcal{H}$ in itself is a feasible solution to $\mathcal{MC}3$ (i.e. (8)).

A.1.5 Affine image matching as a multi-output regression

For an image pair related by the affine map (17), each correspondence $(\tilde{x}_i, \tilde{x}'_i)$ produces a two-dimensional residual

$$r_i(\beta_1, \beta_2) = \left(x'_i - \beta_1^\top \tilde{x}_i,\ y'_i - \beta_2^\top \tilde{x}_i\right) \in \mathrm{R}^2,$$

where $\beta_1, \beta_2 \in \mathrm{R}^3$ are the two active rows of A. Measured in the ℓ_∞ norm, correspondence i is an inlier exactly when

$$\|r_i\|_\infty \leq \epsilon \iff |x'_i - \beta_1^\top \tilde{x}_i| \leq \epsilon \text{ and } |y'_i - \beta_2^\top \tilde{x}_i| \leq \epsilon.$$

Introducing one binary indicator z_i per correspondence and relaxing each coordinate constraint with the big-M term $z_i\mathrm{M}$, we get:

$$\min_{\beta_1, \beta_2, z} \quad \sum_i z_i \tag{19a}$$

$$\text{s.t.} \quad |x'_i - \beta_1^\top \tilde{x}_i| \leq \epsilon + z_i\mathrm{M}, \quad \forall\, i \in \{1, \ldots, \mathrm{N}\} \tag{19b}$$

$$|y'_i - \beta_2^\top \tilde{x}_i| \leq \epsilon + z_i\mathrm{M}, \quad \forall\, i \in \{1, \ldots, \mathrm{N}\} \tag{19c}$$

$$z_i \in \{0, 1\} \quad \forall\, i \in \{1, \ldots, \mathrm{N}\}. \tag{19d}$$

Since both coordinates share the same z_i, a correspondence is kept or discarded as a unit, exactly as one row of (1).

The two constraint families in (19) act on disjoint parameter blocks: the first (19b) only on β_1, the second (19c) only on β_2, coupled solely through the shared binary variables z_i and the common objective $\sum_i z_i$. We exploit this to reduce (19) to the set-hitting form $\mathcal{MC}3$ of Theorem 2.2 at the *per-coordinate* dimension $d = 3$, rather than at the nominal six parameters of β_1 and β_2 combined.

We treat each coordinate as a separate MAXCON instance (1) defined on the common index set $\mathcal{S} = \{1, \ldots, \mathrm{N}\}$ (one index per correspondence) with shared binary variables z_i:

$$\mathcal{D}^x = \{(\tilde{x}_i,\, x_i')\}_{i\in\mathcal{S}}, \qquad \mathcal{D}^y = \{(\tilde{x}_i,\, y_i')\}_{i\in\mathcal{S}},$$

with regressors $\beta_1, \beta_2 \in \mathrm{R}^3$. Since β_1 and β_2 are independent, a subset of correspondences is jointly in consensus *iff* it is in consensus in each coordinate separately, so Theorem 2.2 applies to $\mathcal{D}^x$ and $\mathcal{D}^y$ independently.

Applying it over the $(d+1) = 4$-subsets of $\mathcal{S}$ yields two universal infeasible collections, using the feasibility operator $\mathcal{F}_\epsilon(\cdot)$ on the coordinate-specific data subsets $\mathcal{D}^x_{\mathcal{S}^{d+1}}$ and $\mathcal{D}^y_{\mathcal{S}^{d+1}}$,

$$\mathcal{U}^{d+1}_{\mathcal{I},x} = \{\mathcal{D}^x_{\mathcal{S}^{d+1}} : \mathcal{F}_\epsilon(\mathcal{D}^x_{\mathcal{S}^{d+1}}) = \phi\}, \quad \mathcal{U}^{d+1}_{\mathcal{I},y} = \{\mathcal{D}^y_{\mathcal{S}^{d+1}} : \mathcal{F}_\epsilon(\mathcal{D}^y_{\mathcal{S}^{d+1}}) = \phi\}.$$

Imposing the set-hitting constraints of both collections on the shared binary variables z_i gives the integer program

$$\begin{aligned}
\min_{z} \quad & \sum_{i=1}^{\mathrm{N}} z_i \\
\text{s.t.} \quad & \sum_{i\in\mathcal{S}^{d+1}} z_i \geq 1 \quad \forall\, \mathcal{D}^x_{\mathcal{S}^{d+1}} \in \mathcal{U}^{d+1}_{\mathcal{I},x}, && \text{(20a)} \\
& \sum_{i\in\mathcal{S}^{d+1}} z_i \geq 1 \quad \forall\, \mathcal{D}^y_{\mathcal{S}^{d+1}} \in \mathcal{U}^{d+1}_{\mathcal{I},y}, && \text{(20b)} \\
& z_i \in \{0,1\} \qquad \forall\, i \in \{1, \ldots, \mathrm{N}\}.
\end{aligned}$$

For each $(d+1)$-subset $\mathcal{D}_{\mathcal{S}^{d+1}}$, if *either* coordinate is infeasible — $\mathcal{F}_\epsilon(\mathcal{D}^x_{\mathcal{S}^{d+1}}) = \phi$ or $\mathcal{F}_\epsilon(\mathcal{D}^y_{\mathcal{S}^{d+1}}) = \phi$, i.e. the subset is not jointly in consensus — it contributes the identical set-hitting constraint $\sum_{i\in\mathcal{S}^{d+1}} z_i \geq 1$ on the shared binary variables z_i. The two families can therefore be collapsed onto a single universal infeasible set defined over both coordinates together,

$$\mathcal{U}^{d+1}_{\mathcal{I}} \;:=\; \left\{\, \mathcal{D}_{\mathcal{S}^{d+1}} \;:\; \mathcal{F}_\epsilon(\mathcal{D}^x_{\mathcal{S}^{d+1}}) = \phi \;\text{ or }\; \mathcal{F}_\epsilon(\mathcal{D}^y_{\mathcal{S}^{d+1}}) = \phi \,\right\} \;=\; \mathcal{U}^{d+1}_{\mathcal{I},x} \;\cup\; \mathcal{U}^{d+1}_{\mathcal{I},y}.$$

The two constraint families (20a) and (20b) then merge into a single set-hitting program over $\mathcal{U}^{d+1}_{\mathcal{I}}$,

$$\begin{aligned}
\min_{z} \quad & \sum_{i=1}^{\mathrm{N}} z_i \\
\text{s.t.} \quad & \sum_{i\in\mathcal{S}^{d+1}} z_i \geq 1 \quad \forall\, \mathcal{D}_{\mathcal{S}^{d+1}} \in \mathcal{U}^{d+1}_{\mathcal{I}}, && \text{(21)} \\
& z_i \in \{0,1\} \qquad \forall\, i \in \{1, \ldots, \mathrm{N}\},
\end{aligned}$$

which is exactly $\mathcal{MC}3$ (8) over $\mathcal{U}^{d+1}_{\mathcal{I}}$ with $d = 3$.

A.2 Detailed individual real dataset results

Tables 4–9 give per-setup results for all six datasets. A superscript $*$ on the Algo. 3 runtime marks the setups that Algo. 3 solved to certified optimality *without* using a MILP solver—the set-hitting reduction alone closed the problem. **This is observed in** 58 **of the** 103 **setups, i.e. around** 56% **of setups are solved without Gurobi (**$10/40$ **affine image matching,** $48/63$ **affine epipolar geometry).**

Table 4: Per-setup results on the `boat` dataset. Columns 1–3 give the image frame pair, the number of putative correspondences N, and the outlier % (certified by Algo. 3); columns 4 and 5 are the runtimes of Algo. 3 and $\mathcal{MC}1$ respectively; column 6 is the runtime gain $\mathcal{MC}1/$Algo. 3; columns 7 and 8 are the two methods' final optimality gaps, where the gap is the relative MIP gap (incumbent − bound)/incumbent at termination; and the final column is the inlier gain of Algo. 3 over the starting lower bound $n_{\text{il}}^{\text{lb}}$ (random sampling for 50k iterations).

Pair	N	outlier %	Algo. 3 (s)	$\mathcal{MC}1$ (s)	Gain	Gap		Inlier Gain
						Algo. 3	$\mathcal{MC}1$	
2→6	240	69%	247.1	1.9×10^3	7.8×	0.0%	0.0%	6
1→6	231	65%	208.9*	4.0×10^2	1.9×	0.0%	0.0%	5
3→6	295	63%	716.9	3.0×10^4	42.5×	0.0%	0.0%	7
1→5	357	57%	1430.9	3.8×10^4	26.8×	0.0%	0.0%	7
2→5	341	53%	1536.9	7.5×10^4	48.9×	0.0%	0.0%	8
3→6	208	52%	317.6	2.0×10^3	6.3×	0.0%	0.0%	7
Avg.	279	60%	743.0	2.5×10^4	**22.4**×	0.0%	0.0%	6.7

Table 5: Per-setup results on the `graf` dataset. Columns 1–3 give the image frame pair, the number of putative correspondences N, and the outlier % (certified by Algo. 3); columns 4 and 5 are the runtimes of Algo. 3 and $\mathcal{MC}1$ respectively; column 6 is the runtime gain $\mathcal{MC}1/$Algo. 3; columns 7 and 8 are the two methods' final optimality gaps, where the gap is the relative MIP gap (incumbent − bound)/incumbent at termination; and the final column is the inlier gain of Algo. 3 over the starting lower bound $n_{\text{il}}^{\text{lb}}$ (random sampling for 50k iterations). For the two setups Algo. 3 did not solve within its budget, we instead report its best incumbent (feasible) solution: the Algo. 3 outlier rate is an upper bound ($\leq$).

Pair	N	outlier %	Algo. 3 (s)	$\mathcal{MC}1$ (s)	Gain	Gap		Inlier Gain
						Algo. 3	$\mathcal{MC}1$	
2→4	215	80%	168.4*	8.1×10^4	≥480.3×	0.0%	19.8% (oom)	9
3→5	194	78%	318.1	5.9×10^4	≥185.9×	0.0%	15.8% (oom)	7
5→6	330	≤74%	46834.6	1.2×10^5	≥2.6×	11.2% (oom)	41.9% (oom)	7
2→3	300	74%	18563.8	1.2×10^5	≥6.5×	0.0%	39.6% (oom)	9
2→3	396	73%	31488.1	1.7×10^5	≥5.5×	0.0%	42.6% (timeout)	11
5→6	235	72%	7303.4	4.8×10^4	≥6.6×	0.0%	24.7% (oom)	9
2→3	193	72%	428.5	6.7×10^4	≥156.0×	0.0%	27.3% (oom)	9
3→4	163	≤65%	59365.0	5.1×10^4	≥0.9×	10.2% (timeout)	12.3% (oom)	5
1→2	165	67%	94.3*	1.4×10^5	1509.3×	0.0%	0.0%	7
1→2	302	67%	2451.4	9.2×10^4	≥37.4×	0.0%	25.1% (oom)	11
3→4	275	66%	6622.4	6.8×10^4	≥10.3×	0.0%	27.9% (oom)	8
4→5	327	65%	12234.4	1.1×10^5	≥9.3×	0.0%	34.7% (oom)	10
4→5	244	65%	2627.1	8.9×10^4	≥33.7×	0.0%	28.7% (oom)	9
4→5	161	60%	429.4	9.0×10^4	≥209.4×	0.0%	9.3% (oom)	8
Avg.	250	70%	13494.9	9.4×10^4	**189.5**×	1.5%	25.0%	8.5

Table 6: Per-setup results on the `wall` dataset. Columns 1–3 give the image frame pair, the number of putative correspondences N, and the outlier % (certified by Algo. 3); columns 4 and 5 are the runtimes of Algo. 3 and $\mathcal{MC}1$ respectively; column 6 is the runtime gain $\mathcal{MC}1/\text{Algo. 3}$; columns 7 and 8 are the two methods' final optimality gaps, where the gap is the relative MIP gap $(\text{incumbent} - \text{bound})/\text{incumbent}$ at termination; and the final column is the inlier gain of Algo. 3 over the starting lower bound $n_{\text{il}}^{\text{lb}}$ (random sampling for 50k iterations).

Pair	N	outlier %	Algo. 3 (s)	$\mathcal{MC}1$ (s)	Gain	Gap		Inlier Gain
						Algo. 3	$\mathcal{MC}1$	
1→4	306	83%	1388.5	1.1×10^5	≥81.8×	0.0%	39.0% (oom)	7
2→5	380	83%	6797.8	1.4×10^5	≥20.6×	0.0%	42.4% (oom)	9
4→6	271	82%	345.7*	6.9×10^4	≥198.9×	0.0%	30.0% (oom)	6
1→4	239	81%	350.2	8.0×10^4	≥229.1×	0.0%	31.6% (oom)	7
2→5	307	80%	3469.2	1.3×10^5	≥37.3×	0.0%	38.2% (oom)	8
1→3	339	79%	1248.3*	5.7×10^4	≥46.0×	0.0%	29.7% (oom)	6
5→6	373	79%	3247.1*	1.1×10^5	≥35.3×	0.0%	41.7% (oom)	5
4→6	190	79%	93.1*	6.2×10^4	≥661.3×	0.0%	19.3% (oom)	6
1→3	236	78%	246.5*	3.7×10^4	≥151.5×	0.0%	20.7% (oom)	2
3→5	312	78%	4004.2	5.8×10^4	≥14.4×	0.0%	32.1% (oom)	8
2→5	231	78%	440.8*	4.8×10^4	≥109.6×	0.0%	18.9% (oom)	8
1→4	157	77%	51.6	9.7×10^4	1873.4×	0.0%	0.0%	4
2→5	160	75%	95.2	5.1×10^4	534.6×	0.0%	0.0%	3
3→5	217	75%	327.7	5.2×10^4	≥159.1×	0.0%	19.1% (oom)	6
5→6	258	74%	527.8*	7.9×10^4	≥149.5×	0.0%	32.1% (oom)	7
2→4	358	73%	5833.6	7.4×10^4	≥12.7×	0.0%	32.3% (oom)	8
1→2	261	72%	876.1	5.9×10^4	≥66.9×	0.0%	24.6% (oom)	7
4→5	233	70%	808.2	5.2×10^4	≥64.6×	0.0%	21.8% (oom)	8
2→4	217	64%	452.4	8.1×10^4	≥178.9×	0.0%	8.6% (oom)	6
3→4	361	60%	6039.9	1.1×10^5	≥17.6×	0.0%	25.2% (oom)	11
Avg.	270	76%	1832.2	7.8×10^4	**232.2**×	0.0%	25.4%	6.6

Table 7: Per-setup results on the `dino` dataset. Columns 1–3 give the image frame pair, the number of putative correspondences N, and the outlier % (certified by Algo. 3); columns 4 and 5 are the runtimes of Algo. 3 and $\mathcal{MC}1$ respectively; column 6 is the runtime gain $\mathcal{MC}1/\text{Algo. 3}$; columns 7 and 8 are the two methods' final optimality gaps, where the gap is the relative MIP gap $(\text{incumbent} - \text{bound})/\text{incumbent}$ at termination; and the final column is the inlier gain of Algo. 3 over the starting lower bound $n_{\text{il}}^{\text{lb}}$ (random sampling for 50k iterations).

Pair	N	outlier %	Algo. 3 (s)	$\mathcal{MC}1$ (s)	Gain	Gap		Inlier Gain
						Algo. 3	$\mathcal{MC}1$	
30→33	191	78%	4040.9	2.9×10^4	≥7.3×	0.0%	29.2% (oom)	2
31→34	200	72%	5896.2*	2.6×10^4	≥4.5×	0.0%	25.5% (oom)	5
6→9	168	70%	1715.0	4.7×10^4	≥27.2×	0.0%	31.3% (oom)	2
28→31	125	70%	387.2*	2.2×10^4	56.1×	0.0%	0.0%	4
14→17	170	69%	1745.9*	4.0×10^4	≥23.0×	0.0%	36.5% (oom)	5
27→30	142	68%	592.5*	4.2×10^4	≥71.6×	0.0%	26.2% (oom)	0
5→8	147	64%	931.9*	5.1×10^4	≥55.2×	0.0%	16.1% (oom)	4
15→18	144	62%	813.2*	4.0×10^4	≥49.6×	0.0%	20.2% (oom)	4
13→16	127	62%	364.5*	4.3×10^4	≥117.9×	0.0%	13.1% (oom)	2
8→11	156	60%	1161.2*	4.5×10^4	≥39.1×	0.0%	19.4% (oom)	2
4→7	193	60%	3286.1	8.2×10^4	≥25.0×	0.0%	28.0% (oom)	3
7→10	165	59%	1410.7*	6.8×10^4	≥48.0×	0.0%	18.5% (oom)	3
29→31	174	59%	2194.2*	5.9×10^4	≥26.8×	0.0%	30.2% (oom)	3
26→29	170	57%	1654.9*	5.1×10^4	≥31.1×	0.0%	20.4% (oom)	1
11→14	129	55%	431.6*	1.4×10^5	330.2×	0.0%	0.0%	2
16→19	176	53%	2481.1*	1.7×10^5	≥69.9×	0.0%	3.2% (timeout)	3
9→12	163	53%	1549.7*	1.7×10^5	≥112.0×	0.0%	3.4% (timeout)	4
12→15	128	52%	423.5*	2.4×10^4	57.0×	0.0%	0.0%	0
10→13	127	52%	450.2	8.2×10^3	18.3×	0.0%	0.0%	1
28→30	181	50%	2407.7*	7.1×10^4	≥29.6×	0.0%	20.5% (oom)	0
32→35	158	50%	1276.5	6.7×10^4	52.5×	0.0%	0.0%	0
17→20	202	48%	4895.4*	1.7×10^5	≥35.4×	0.0%	6.1% (timeout)	6
30→32	160	48%	1439.1*	1.2×10^5	82.5×	0.0%	0.0%	2
31→33	185	45%	3136.1*	1.1×10^5	34.7×	0.0%	0.0%	2
25→28	190	44%	3396.7*	3.5×10^4	10.4×	0.0%	0.0%	1
14→16	169	43%	1973.8*	1.4×10^4	6.9×	0.0%	0.0%	1
Avg.	163	58%	1925.2	6.8×10^4	**54.7**×	0.0%	13.4%	2.4

Table 8: Per-setup results on the `kapel` dataset. Columns 1–3 give the image frame pair, the number of putative correspondences N, and the outlier % (certified by Algo. 3); columns 4 and 5 are the runtimes of Algo. 3 and $\mathcal{MC}1$ respectively; column 6 is the runtime gain $\mathcal{MC}1/\text{Algo. 3}$; columns 7 and 8 are the two methods' final optimality gaps, where the gap is the relative MIP gap $(\text{incumbent} - \text{bound})/\text{incumbent}$ at termination; and the final column is the inlier gain of Algo. 3 over the starting lower bound $n_{\text{il}}^{\text{lb}}$ (random sampling for 50k iterations).

Pair	N	outlier %	Algo. 3 (s)	$\mathcal{MC}1$ (s)	Gain	Gap		Inlier Gain
						Algo. 3	$\mathcal{MC}1$	
15→25	156	67%	1329.3*	4.2×10^4	≥31.6×	0.0%	27.9% (oom)	2
25→64	134	63%	492.6*	3.5×10^4	70.4×	0.0%	0.0%	2
18→64	167	62%	1528.5*	7.9×10^4	51.5×	0.0%	0.0%	3
44→58	191	58%	5846.7	5.5×10^4	≥9.5×	0.0%	30.0% (oom)	3
47→75	133	52%	845.0	1.1×10^5	≥127.0×	0.0%	7.9% (oom)	2
20→66	129	51%	577.6*	4.0×10^3	6.9×	0.0%	0.0%	3
36→47	197	50%	8053.3	8.9×10^4	≥11.0×	0.0%	25.3% (oom)	4
48→74	130	50%	1102.3	1.4×10^2	0.1×	0.0%	0.0%	3
74→77	133	49%	731.5*	1.4×10^2	0.2×	0.0%	0.0%	1
23→67	132	48%	643.8*	1.3×10^3	2.0×	0.0%	0.0%	4
38→77	126	48%	682.1*	7.2×10^3	10.6×	0.0%	0.0%	3
38→73	175	47%	4782.5*	7.0×10^4	≥14.7×	0.0%	20.3% (oom)	3
35→48	163	47%	2393.1*	7.0×10^4	≥29.3×	0.0%	12.8% (oom)	3
20→24	155	46%	1690.6	1.3×10^2	0.1×	0.0%	0.0%	3
22→24	136	46%	935.7*	9.5×10^3	10.2×	0.0%	0.0%	4
14→18	146	44%	1698.6*	1.3×10^3	0.8×	0.0%	0.0%	5
13→18	137	44%	1552.3	1.4×10^3	0.9×	0.0%	0.0%	3
37→47	189	41%	5680.2*	1.7×10^5	≥30.5×	0.0%	3.9% (timeout)	3
Avg.	152	51%	2253.7	4.1×10^4	**22.6**×	0.0%	7.1%	3.0

Table 9: Per-setup results on the `valbonne` dataset. Columns 1–3 give the image frame pair, the number of putative correspondences N, and the outlier % (certified by Algo. 3); columns 4 and 5 are the runtimes of Algo. 3 and $\mathcal{MC}1$ respectively; column 6 is the runtime gain $\mathcal{MC}1/\text{Algo. 3}$; columns 7 and 8 are the two methods' final optimality gaps, where the gap is the relative MIP gap $(\text{incumbent} - \text{bound})/\text{incumbent}$ at termination; and the final column is the inlier gain of Algo. 3 over the starting lower bound $n_{\text{il}}^{\text{lb}}$ (random sampling for 50k iterations).

Pair	N	outlier %	Algo. 3 (s)	$\mathcal{MC}1$ (s)	Gain	Gap		Inlier Gain
						Algo. 3	$\mathcal{MC}1$	
11→14	177	69%	3236.9*	3.6×10^4	≥11.0×	0.0%	36.0% (oom)	2
6→13	126	65%	392.8*	9.2×10^4	234.9×	0.0%	0.0%	1
3→7	147	58%	950.0*	3.7×10^4	39.4×	0.0%	0.0%	3
4→11	169	58%	2310.5	4.9×10^4	≥21.2×	0.0%	23.1% (oom)	3
0→5	171	57%	2133.4*	4.1×10^2	0.2×	0.0%	0.0%	2
4→7	133	56%	613.5*	1.9×10^4	30.4×	0.0%	0.0%	3
0→6	161	55%	2118.1*	6.1×10^4	≥28.6×	0.0%	8.9% (oom)	2
2→11	127	53%	574.7*	1.2×10^3	2.1×	0.0%	0.0%	4
6→10	158	52%	1625.3*	8.7×10^4	≥53.7×	0.0%	10.9% (oom)	1
5→11	198	52%	6324.9*	9.1×10^4	≥14.4×	0.0%	27.7% (oom)	2
2→8	188	50%	4636.1*	9.2×10^4	≥19.8×	0.0%	18.8% (oom)	3
10→14	179	50%	3304.4*	8.0×10^4	≥24.1×	0.0%	20.5% (oom)	3
5→12	154	47%	2141.7	1.6×10^4	7.7×	0.0%	0.0%	4
3→11	162	47%	4655.6	8.3×10^4	≥17.9×	0.0%	7.8% (oom)	3
0→4	136	46%	1103.4*	8.3×10^2	0.7×	0.0%	0.0%	3
8→10	161	46%	1922.6*	5.7×10^4	29.5×	0.0%	0.0%	1
9→13	163	45%	3436.2	1.7×10^5	≥50.4×	0.0%	2.7% (timeout)	5
3→8	184	44%	4665.8*	1.7×10^5	≥37.2×	0.0%	5.0% (timeout)	3
1→8	125	42%	650.7*	5.6×10^2	0.9×	0.0%	0.0%	2
Avg.	159	52%	2463.0	6.1×10^4	**32.9**×	0.0%	8.5%	2.6