Consensus Maximization with Linear Matrix Inequality Constraints

Pablo Speciale[1], Danda P. Paudel[2], Martin R. Oswald[1],
Till Kroeger[2], Luc V. Gool[2,4], and Marc Pollefeys[1,3]

[1] Department of Computer Science, ETH Zürich.
[3] Microsoft, Redmond, USA
{pablo, moswald, marc.pollefeys}@inf.ethz.ch

[2] Computer Vision Laboratory, D-ITET, ETH Zürich
[4] VISICS, ESAT/PSI, KU Leuven, Belgium
{paudel, kroegert, vangool}@vision.ee.ethz.ch

Abstract

Consensus maximization has proven to be a useful tool for robust estimation. While randomized methods like RANSAC are fast, they do not guarantee global optimality and fail to manage large amounts of outliers. On the other hand, global methods are commonly slow because they do not exploit the structure of the problem at hand.

In this paper, we show that the solution space can be reduced by introducing Linear Matrix Inequality (LMI) constraints. This leads to significant speed ups of the optimization time even for large amounts of outliers, while maintaining global optimality. We study several cases in which the objective variables have a special structure, such as rotation, scaled-rotation, and essential matrices, which are posed as LMI constraints. This is very useful in several standard computer vision problems, such as estimating Similarity Transformations, Absolute Poses, and Relative Poses, for which we obtain compelling results on both synthetic and real datasets. With up to 90 percent outlier rate, where RANSAC often fails, our constrained approach is consistently faster than the non-constrained one - while finding the same global solution.

1. Introduction

One of the major difficulties of many central computer vision problems – besides the proper handling of noise and incomplete data – is the robust detection of outliers. For many optimization methods, the number of outliers has a tremendous impact on the runtime or even on the solvability. A common approach for robust estimation is, therefore, to explicitly maximize the number of inliers for a given problem – also called *Consensus Maximization*. A large number of optimization methods for robust estimation have been proposed in the literature which can be roughly divided into local and global optimization methods.

Acknowledgements. This project has received funding from the European Unions Horizon 2020 research and innovation programme, and European Research Council project VarCity, under grant agreements No. 637221 and 273940, respectively.

Local Optimization Methods. With currently more than 15K citations, RANSAC [13], is by far the most popular method. It has been used in numerous applications and many extensions have been proposed, e.g. [8, 33, 34] (see [7, 28] for an overview). The great advantage is its simplicity and effectiveness for various scenarios, but it also has a number of shortcomings: 1) it does not guarantee optimality and only finds a local optimum, 2) it cannot find the exact solution if it is not contained in the sampling set, and 3) its expected computation time grows exponentially with large amounts of outliers.

Global Optimization Methods. Global methods commonly have considerably larger computational costs, as they are mostly based on exhaustive search within the entire optimization domain. Almost every global method uses the **Branch-and-Bound (BnB)** strategy to make the search tractable, e.g. [1, 2, 16, 22, 38]. Similar to our approach, several methods use **Mixed Integer Programming (MIP)** [4, 9, 22, 36] within the BnB optimization in order to solve the overall problem faster. Recently [5] proposed to cast the problem of consensus maximization as a tree search problem which is then traversed with **A*-search** for faster optimization. This method does not need linearization of the residual and only traverses a small subset of the tree compared to exhaustive methods like [10, 26].

Application-wise, many related works are specialized to a particular problem class, like linear problems [22], pseudo-convex problems [5, 21], or, they are even more specialized to a specific type of geometric problem, for instance problems including rotations [2, 16], rotation+focal length [1], translation [14], rotation+translation+scale [27] or essential matrices [37]. Most of these methods specialize on a particular problem and their application to a different problem class is not necessarily straightforward.

In this paper, we propose a general optimization framework that covers all problems that can be expressed with LMI constraints and, therefore, tackles the majority of the aforementioned problem classes.

Contributions. We argue that many computer vision problems have a special structure that can be leveraged in global robust estimation methods to make them much more competitive compared to local randomized approaches.

- In particular, we propose general LMI constraints which have the potential to be used in a variety of geometric problems. The derived constraints include: rigid-body, rigid-body + scale, restricted rotations, and essential matrix.
- These LMI constraints are used within the BnB paradigm to optimally solve the consensus maximization problem for: similarity transformation, absolute pose, and relative pose estimation.
- We show that the usage of LMI constraints speeds up the search process. Further observation (for the absolute pose problem) shows that the addition of stronger LMI constraints (e.g. cameras must be within a certain angle) makes the optimization process even faster.

Paper Outline. We first introduce notation and background theory (Sec. 2), before we state the problem (Sec. 3). We then define the transformation form considered in this paper (Sec. 4), and derive several LMI constraints for various geometric constraints (Sec. 5). After that, we discuss (Sec. 6) and experimentally evaluate (Sec. 7) three particular computer vision problems. Finally, we conclude (Sec. 8) the paper and state potential future work.

2. Notation & Background

This section introduces the notations that we will be using throughout the paper. Algebraic definitions and a geometric interpretation of LMI constraints are introduced and discussed. Finally, we discuss the analogy between LMI constraint quadratic and semi-definite programming.

2.1. Positive Semi-definite Matrix Formulations

When dealing with matrices, $\mathsf{A} \succeq 0$ (resp. $\mathsf{A} \succ 0$) means that the symmetric matrix A is positive semi-definite (resp. positive-definite). $\mathsf{A} = (a_{ij})$ is the element-wise representation of a $m \times n$ matrix. Its row-wise representation is $\mathsf{A} = [\mathsf{a}_1, \dots, \mathsf{a}_i, \dots, \mathsf{a}_m]^\intercal$, where a_i is a n-dimensional vector. For a given set of symmetric matrices $\mathcal{K} = \{\mathsf{A}_i\}_{i=1}^{l}$, $\mathcal{K} \succeq 0$ implies $\mathsf{A}_i \succeq 0$ for all $i = 1, \dots, l$. It is important to note that $\mathcal{K} \succeq 0$ is equivalent to $\mathsf{A} = \mathrm{diag}(\mathsf{A}_1, \mathsf{A}_2, \dots, \mathsf{A}_l) \succeq 0$.

2.2. Spectrahedron, LMI, and SDP

Definition 2.1 (Spectrahedron [35]) *A spectrahedron is the intersection of positive semi-definite matrices with an affine-linear space. An n-dimensional affine-linear space of a real symmetric matrix can be parametrized by* $\mathsf{A}(\mathsf{y}) = \mathsf{A}_0 + \sum_i^n y_i \mathsf{A}_\mathsf{i}$, *for* $\mathsf{y} = [y_1,\ y_2,\ \dots, y_n]^\intercal \in \mathbb{R}^\mathrm{n}$. *Therefore, a spectrahedron can be defined by a set* $\mathcal{S}$:

$$\mathcal{S} = \{\mathsf{y} \in \mathbb{R}^\mathrm{n} : \mathsf{A}(\mathsf{y}) \succeq 0\}. \tag{1}$$

A **Linear Matrix Inequality (LMI)** is the constraint on $\mathsf{y} \in \mathbb{R}^\mathrm{n}$ such that $\mathsf{A}(\mathsf{y}) \succeq 0$. A **Semi-definite Program (SDP)** consists of minimizing (or maximizing) a linear objective subject to LMI constraints. It is a convex optimization problem that can be efficiently solved using interior-point methods [3].

2.3. LMI Constrained Quadratic Programming

A LMI constrained quadratic programming is a problem of the form:

$$\begin{aligned} \underset{\mathsf{y}}{\text{minimize}} \quad & \mathsf{y}^\intercal \mathsf{Q} \mathsf{y} + \mathsf{q}^\intercal \mathsf{y} + r, \\ \text{subject to} \quad & \mathsf{A}(\mathsf{y}) \succeq 0, \end{aligned} \tag{2}$$

where Q is a real symmetric matrix, q a real vector, and r a real scalar.

For $\mathsf{Q} \succeq 0$, the problem of (2) can be optimally solved using SDP. Note that $\mathsf{Q} \succeq 0$ can always be factorized into $\mathsf{Q} = \mathsf{M}^\intercal \mathsf{M}$ for some matrix M, using the Cholesky decomposition. Therefore (2) is equivalent to the following SDP:

$$\begin{aligned} \underset{\mathsf{y},\theta}{\text{minimize}} \quad & \theta, \\ \text{subject to} \quad & \begin{pmatrix} \mathsf{I} & \mathsf{M}\mathsf{y} \\ \mathsf{y}^\intercal \mathsf{M}^\intercal & \theta - \mathsf{q}^\intercal \mathsf{y} - r \end{pmatrix} \succeq 0, \\ & \mathsf{A}(\mathsf{y}) \succeq 0. \end{aligned} \tag{3}$$

3. Consensus Maximization with LMI

Consider a geometric transformation $T(\mathsf{x}) : U \to V$ that relates a pair of measurements $\mathcal{P} = \{U, V\}$. Let $\gamma(\mathsf{x})$ be the residual error for a known $\mathcal{P}$ and the estimate x. The problem of maximizing the measurements' consensus (i.e. inlier set) under a LMI constraint is,

Problem 3.1 *Given a set of measurement pairs* $\mathcal{Z} = \{\mathcal{P}_i\}_{i=1}^{n}$ *and a threshold* ϵ,

$$\begin{aligned} \underset{\mathsf{x}, \zeta \subseteq \mathcal{Z}}{\text{maximize}} \quad & |\zeta|, \\ \text{subject to} \quad & \gamma_i(\mathsf{x}) \leq \epsilon, \quad \forall \mathcal{P}_i \in \zeta, \\ & \mathsf{A}(\mathsf{x}) \succeq 0. \end{aligned} \tag{4}$$

In general, solving (4) exactly is non-trivial, as this is a NP-hard combinatorial optimization problem. Such problems are usually solved using sample-and-test techniques, such as RANSAC, with no guarantee on the optimality of the results. In contrast, exact methods are based on variations of tree search algorithms [5, 10, 22, 26, 38]. Our following proposition is concerned about the optimal solution search for a class of such problems.

Proposition 3.2 (Consensus with LMIs) *Problem 3.1 can be solved optimally using a tree search method for linear* $\gamma_i(\mathsf{x})$ *or quadratic residuals* $\gamma_i(\mathsf{x}) = \mathsf{x}^\intercal \mathsf{Q}_i \mathsf{x} + \mathsf{q}_i^\intercal \mathsf{x} + r_i$ *with* $\mathsf{Q}_i \succeq 0$.

Proof It is sufficient to show that the problem of minimizing the maximum of $\gamma_i(\mathsf{x})$ under a LMI constraint, as given below, is convex.

$$\begin{aligned} \underset{\mathsf{x}}{\text{minimize}} \quad & \underset{i}{\text{maximum}} \qquad \gamma_i(\mathsf{x}), \\ \text{subject to} \quad & \qquad\qquad \mathsf{A}(\mathsf{x}) \succeq 0. \end{aligned} \tag{5}$$

For linear $\gamma_i(\mathsf{x})$, (5) is equivalent to the following SDP,

$$\begin{aligned} \underset{\mathsf{x},\theta}{\text{minimize}} \quad & \theta, \\ \text{subject to} \quad & \gamma_i(\mathsf{x}) \leq \theta, \quad \forall i, \\ & \mathsf{A}(\mathsf{x}) \succeq 0. \end{aligned} \tag{6}$$

Similarly, if $\gamma_i(\mathsf{x}) = \mathsf{x}^\intercal \mathsf{Q}_i \mathsf{x} + \mathsf{q}_i^\intercal \mathsf{x} + r_i$ with $\mathsf{Q}_i \succeq 0$, then (5) can be solved using the following SDP, for $\mathsf{Q}_i = \mathsf{M}_i^\intercal \mathsf{M}_i$,

$$\begin{aligned} \underset{\mathsf{x},\theta}{\text{minimize}} \quad & \theta, \\ \text{subject to} \quad & \begin{pmatrix} \mathsf{I} & \mathsf{M}_i \mathsf{x} \\ \mathsf{x}^\intercal \mathsf{M}_i^\intercal & \theta - \mathsf{q}_i^\intercal \mathsf{x} - r_i \end{pmatrix} \succeq 0, \quad \forall i, \\ & \mathsf{A}(\mathsf{x}) \qquad\qquad\qquad\qquad \succeq 0. \end{aligned} \tag{7}$$

Alternatively, following a similar argument of convexity as in [11], one can show that (5) is a LP-type [24] (or generalized linear program). This concludes the proof. ∎

Mixed Integer Programming. Before entering into further details, we discuss the choice of solving (4) using Mixed Integer Semi-Definite Programming (MI-SDP) [9, 36]. MI-SDP framework can solve (4) optimally, which can be restated as:

$$\begin{aligned} \underset{\mathsf{x},\mathsf{z}}{\text{minimize}} \quad & \sum_i z_i, \\ \text{subject to} \quad & \gamma_i(\mathsf{x}) \leq \epsilon + z_i \, \mathcal{M}, \quad \forall i, \\ & z_i \in \{0, 1\}, \\ & \mathsf{A}(\mathsf{x}) \succeq 0. \end{aligned} \tag{8}$$

where $\mathsf{z} = \{z_i\}_{i=1}^n$ are binary variables and $\mathcal{M}$ is *a large enough* positive constant. It is a common practice in optimization to ignore constraints by using a constant such as $\mathcal{M}$. See [6, Ch. 7] for guidelines on selecting this constant. Intuitively, the data pair generating the residual $\gamma_i(x)$ will be considered as an outlier if $z_i = 1$. Therefore, for the given optimal solution z^* to (8), the maximum consensus set can be obtained by,

$$\zeta^* = \{\, i \,|\, z_i^* = 0\}. \tag{9}$$

4. Transformation Equation

Although Proposition 3.2 suggests that the general class of problems (4) can be optimally solved, in this work we are concerned with problems of the following form,

$$\beta_i(\mathsf{x}) \, \mathsf{v}_i = \mathsf{B}_i(\mathsf{x}) \, \mathsf{u}_i + \mathsf{b}(\mathsf{x}), \tag{10}$$

where $\{\mathsf{u}_i, \mathsf{v}_i\}$ are the measurement vectors of the pair $\mathcal{P}_i$, and $\mathsf{B}_i(\mathsf{x})$, $\mathsf{b}(\mathsf{x})$, and $\beta_i(\mathsf{x})$ are the terms linear on x which form the transformation $T(\mathsf{x})$. For a given problem, we wish to enforce the structural constraint of $T(\mathsf{x})$ in terms of LMIs, while minimizing the residual error of (10).

4.1. Residual with Noise Model

We model the noise as a Gaussian process. The dissimilarity measure between two corresponding measurements is therefore expressed in terms of the *generalized squared interpoint distance* (also known as squared Mahalanobis distance). For a given pair $\mathcal{P}_i$, the residual error (or dissimilarity measure) is given by,

$$\begin{aligned} \gamma_i(\mathsf{x}) &= \Delta_i(\mathsf{x})^\intercal \, \Sigma^{-1} \, \Delta_i(\mathsf{x}), \\ \Delta_i(\mathsf{x}) &= \mathsf{B}_i(\mathsf{x}) \, \mathsf{u}_i + \mathsf{b}(\mathsf{x}) - \beta_i(\mathsf{x}) \, \mathsf{v}_i \end{aligned} \tag{11}$$

where Σ is the covariance matrix of a known distribution.

Remark 4.1 *The residual error* (11) *is a quadratic function of the form* $\gamma_i(\mathsf{x}) = \mathsf{x}^\intercal \mathsf{Q}_i \mathsf{x} + \mathsf{q}_i^\intercal \mathsf{x} + r_i$ *with* $\mathsf{Q}_i \succeq 0$.

4.2. Residual Minimization

Once the optimal inlier set ζ for Problem 3.1 is obtained, the best estimate x that minimizes the collective residual errors for all the inliers, while satisfying the LMI constraint, can be obtained using the following result. This can be considered as a refinement step.

Result 4.2 *The optimal estimate* x *that minimizes the sum of residuals for all inlier pairs* $\zeta = \{\mathcal{I}_j\}_{j=1}^m \subseteq \mathcal{Z}$ *can be obtained using the LMI constrained quadratic programming,*

$$\begin{aligned} \underset{\mathsf{x}}{\text{minimize}} \quad & \mathsf{x}^\intercal \Big(\sum_{j=1}^m \mathsf{Q}_j \Big) \mathsf{x} + \Big(\sum_{j=1}^m \mathsf{q}_j \Big)^\intercal \mathsf{x} + \sum_{j=1}^m r_j, \\ \text{subject to} \quad & \mathsf{A}(\mathsf{x}) \succeq 0. \end{aligned} \tag{12}$$

This is a convex problem, that can be solved efficiently using the SDP as discussed in Section 2.3. It follows that, $\mathsf{Q}_j \succeq 0 \implies \sum_{j=1}^m \mathsf{Q}_j \succeq 0$.

5. LMI Constraints

In this section, we introduce four LMI constraints that we will use in our experiments. Two of these constraints were recently proposed in [15, 29]. The other two constraints are presented for the first time in this paper. First, we define the function $\mathcal{L} : \mathbb{R}^{3\times3} \to \mathbb{R}^{4\times4}$ of the form,

$$\mathcal{L}(\mathsf{A}) = \begin{bmatrix} a_{11}+a_{22}+a_{33} & a_{32}-a_{23} & a_{13}-a_{31} & a_{21}-a_{12} \\ a_{32}-a_{23} & a_{11}-a_{22}-a_{33} & a_{21}+a_{12} & a_{13}+a_{31} \\ a_{13}-a_{31} & a_{21}+a_{12} & a_{22}-a_{11}-a_{33} & a_{32}+a_{23} \\ a_{21}-a_{12} & a_{13}+a_{31} & a_{32}+a_{23} & a_{33}-a_{11}-a_{22} \end{bmatrix}. \tag{13}$$

5.1. Orbitope and Rotation Matrix

Definition 5.1 (Orbitope [29]) *An orbitope is the convex hull of an orbit of a compact algebraic group that acts linearly on a real vector space. The orbit has the structure of a real algebraic variety, and the orbitope is a convex semi-algebraic set.*

A 3-dimensional rotation matrix $\mathsf{R} \in SO(3)$ has dimension three. However, its tautological orbitope is a convex body of dimension nine. The following theorem is a key ingredient of this work:

Theorem 5.2 (SO(3) Orbitope [29]) *The tautological orbitope $conv(SO(3))$ is a spectrahedron whose boundary is a quartic hypersurface. A $3{\times}3$ matrix A lies in $conv(SO(3))$ if and only if,*

$$\mathsf{I}_{4\times 4} + \mathcal{L}(\mathsf{A}) \succeq 0. \tag{14}$$

For some applications, the desired rotation matrix must have a restricted rotation angle around an arbitrary rotation axis. For example, the relative rotation between two cameras cannot be too large for them to share a common field of view. Such rotation angle restrictions can be expressed as LMI constraints using the following Lemma.

Lemma 5.3 (Bounded SO(3) [15]) *Consider $\mathsf{R} \in SO(3)$ expressed in the angle-axis form by a rotation angle θ around an arbitrary axis. The eigen-decomposition of the symmetric matrix $\mathsf{R} + \mathsf{R}^\intercal$ is of the form,*

$$\mathsf{R} + \mathsf{R}^\intercal = \mathsf{U} \begin{bmatrix} 2\,cos\,\theta & 0 & 0 \\ 0 & 2\,cos\,\theta & 0 \\ 0 & 0 & 2 \end{bmatrix} \mathsf{U}^{-1}. \tag{15}$$

For $|\theta| \leq 90°$, the symmetric matrix $\mathsf{R} + \mathsf{R}^\intercal \succeq 0$. In fact, the rotation angle is within the given upper bound, $|\theta| \leq \overline{\theta} \leq 90°$, if and only if,

$$\mathsf{R} + \mathsf{R}^\intercal \succeq 2\,cos\,\overline{\theta}\,\mathsf{I}_{3\times 3}. \tag{16}$$

5.2. Transformations Coupled with Rotation

Some transformations involving both rotation and an extra scale (e.g. Metric Transformation) can be expressed with the help of a scaled-rotation matrix. The following definition deals with the structure of a scaled-rotation matrix.

Definition 5.4 (SSO(3)) *Given a real, compact, linear algebraic group $\mathcal{H}$, a 3-dimensional scaled-special orthogonal group is defined by,*

$$SSO(3) = \{\mathsf{H} \in \mathcal{H} : \mathsf{H}\mathsf{H}^\intercal = \alpha^2\,\mathsf{I}_{3\times 3}, det(\mathsf{H}) = \alpha^3, \alpha > 0\}. \tag{17}$$

In the following proposition, we provide a convex relaxation for the scaled-rotation matrix as a LMI.

Proposition 5.5 (SSO(3) and SO(3) Orbitope) *$\forall\,\mathsf{S} \in SSO(3)$ there exists $\mathsf{A} \in conv(SO(3))$ such that $\mathsf{S} = \alpha\,\mathsf{A}$, if and only if $\exists\,\alpha > 0$:*

$$\alpha\,\mathsf{I}_{4\times 4} + \mathcal{L}(\mathsf{S}) \succeq 0. \tag{18}$$

Proof If $\mathsf{S} \in SSO(3) \iff \mathsf{S} = \alpha\,\mathsf{A}$ for $\mathsf{A} \in SO(3)$ and $\alpha > 0$. Notice from (13) that $\frac{1}{\alpha}\mathcal{L}(\mathsf{S}) = \mathcal{L}(\frac{\mathsf{S}}{\alpha})$. Therefore, (18) is equivalent to $\mathsf{I}_{4\times 4} + \mathcal{L}(\frac{\mathsf{S}}{\alpha}) \succeq 0$. After replacing $\mathsf{A} = \frac{\mathsf{S}}{\alpha}$, we get the same result as in Theorem 5.2. The backward proof for equivalence can be obtained in a very similar manner. ∎

Remark 5.6 *For a given $\alpha > 0$ and $\mathsf{S} = \alpha\,\mathsf{R}$ with R bounded by $|\theta| \leq \overline{\theta} \leq 90°$, the following LMI must also hold true,*

$$\mathsf{S} + \mathsf{S}^\intercal \succeq 2\,\alpha\,cos\overline{\theta}\,I_{3\times 3}. \tag{19}$$

In the calibrated relative pose formulation, a rotation matrix appears with a skew-symmetric matrix, the so-called Essential matrix. More formally, the normalized Essential matrix is defined as follows.

Definition 5.7 (Normalized Essential Matrix) *The set of normalized Essential matrices for two cameras related by rotation R and translation t is given by,*

$$\mathcal{E} = \{\mathsf{E} = [\mathsf{t}]_\times \mathsf{R} : \mathsf{R} \in SO(3), \|\mathsf{t}\| = 1\}. \tag{20}$$

where $[\mathsf{t}]_\times$ is a $3{\times}3$ skew-symmetric matrix.

We will now show that the structural constraint of the Essential matrix can also be expressed as LMI using our following proposition.

Proposition 5.8 (Essential Matrix and Orbitope) *A $3{\times}3$ matrix E belongs to the set of normalized Essential matrices, $\mathcal{E}$ of (20), only if,*

$$2\,\mathsf{I}_{4\times 4} + \mathcal{L}(\mathsf{E}) \succeq 0. \tag{21}$$

Proof For any $\mathsf{E} \in \mathcal{E}$, its singular value decomposition is given by $\mathsf{E} = \mathsf{U}\,diag(1,1,0)\,\mathsf{V}^\intercal$. This can further be decomposed into $\mathsf{E} = \mathsf{A} + \mathsf{B}$, with $\mathsf{A} \in SO(3)$ and $\mathsf{B} = \mathsf{U}\,diag(0,0,-|\mathsf{U}\mathsf{V}^\intercal|)\,\mathsf{V}^\intercal$. Using (13), one can show that $\mathcal{L}(\mathsf{B}) \succeq -\mathsf{I}_{4\times 4}$. Furthermore, from (14), $\mathcal{L}(\mathsf{A}) \succeq -\mathsf{I}_{4\times 4}$. Therefore, $\mathcal{L}(\mathsf{E}) = \mathcal{L}(\mathsf{A}) + \mathcal{L}(\mathsf{B}) \succeq -2\,\mathsf{I}_{4\times 4}$, which leads to (21). ∎

6. Multiple View Geometry Problems

In this section, we present three examples of multiple view geometry problems formulated as in Problem 3.1 (Consensus Maximization with LMIs). Different problems specify different variable terms, in reference to (10), and their LMI constraints are summarized in Table 1.

Transformations	Constraints	$\beta_i(\mathsf{x})$	$\mathsf{B}_i(\mathsf{x})$	$\mathsf{b}(\mathsf{x})$	LMIs
Similarity	$\mathsf{S}(\mathsf{x}) \in SSO(3)$	1	$\mathsf{S}(\mathsf{x})$	$\mathsf{t}(\mathsf{x})$	$\mathcal{K}_s \succeq 0$
Absolute Pose	$\mathsf{R}(\mathsf{x}) \in SO(3)$	$\mathsf{r}_3(\mathsf{x})^\intercal \mathsf{u}_i + t_3(\mathsf{x})$	$\mathsf{R}(\mathsf{x})$	$\mathsf{t}(\mathsf{x})$	$\mathcal{K}_a \succeq 0$
Relative Pose	$\mathsf{E}(\mathsf{x}) \in \mathcal{E}$	$\left[(\mathsf{n}_i)_1\mathsf{e}_2(\mathsf{x}) - (\mathsf{n}_i)_2\mathsf{e}_1(\mathsf{x})\right]^\intercal \mathsf{u}_i$	$[\mathsf{n}_i]_\times \mathsf{E}(\mathsf{x})$	0	$\mathcal{K}_r \succeq 0$

Table 1: Summary of the residual terms and constraints for three different example problems. $\beta_i(\mathsf{x})$, $\mathsf{B}_i(\mathsf{x})$, and $\mathsf{b}(\mathsf{x})$, are the variable terms of (10) compared to (22), (24), and (26). The LMI constraints are given in (23), (25), and (27).

6.1. Similarity Transformation

We consider a set of images acquired by a collection of cameras which observe the same scene. These images are then fed into a SfM pipeline [17, 30] to obtain a 3D Reconstruction. Let $\{\mathsf{u}_i, \mathsf{v}_i\} \in \mathbb{R}^3, i = 1, \ldots, n$, with $n \geq 4$, be the Cartesian coordinate vector pairs of the SfM-induced camera centers U_i and their real world positions V_i (e.g. GPS measurements). Since the SfM reconstruction is metric, SfM-induced cameras and their world measurements are related by

$$\mathsf{v}_i = \mathsf{S}(\mathsf{x})\,\mathsf{u}_i + \mathsf{t}(\mathsf{x}), \tag{22}$$

where $\mathsf{S}(\mathsf{x})$ is a $3{\times}3$ scaled-rotation matrix, $\mathsf{t}(\mathsf{x})$ a $3{\times}1$ translation vector, and $\mathsf{x} \in \mathbb{R}^{12}$. Notice that (22) is analogous to (10), hence its residual error can be written as in (11). On the other hand, a direct application of Proposition 5.5 provides a convex relaxation, as a LMI, to the scaled-rotation matrix constraint, i.e. $\mathsf{S}(\mathsf{x}) \in SSO(3)$.

Corollary 6.1 $\mathsf{S}(\mathsf{x}) = \alpha\,\mathsf{A}$ *such that* $\mathsf{A} \in conv(SO(3))$ *is possible if and only if the following LMI is feasible for some* $\mathsf{x} \in \mathbb{R}^{12}$ *and* $\alpha > 0$:

$$\mathcal{K}_s = \left\{\alpha\,\mathsf{I}_{4\times4} + \mathcal{L}(\mathsf{S}(\mathsf{x}))\right\} \succeq 0. \tag{23}$$

6.2. Absolute Pose

We consider measurement vector pairs $\{\mathsf{u}_i, \mathsf{v}_i\} \in \mathbb{R}^3$, $i = 1, \ldots, n$, with $n \geq 5$, where u_i is the Cartesian representation of the scene point U_i in the world frame and v_i is the homogeneous representation of the image point V_i in the camera frame of a calibrated camera. If $[\mathsf{R}|\mathsf{t}]$ is the camera pose w.r.t. the world frame, then the scene and image points are related by,

$$\left(\mathsf{r}_3(\mathsf{x})^\intercal \mathsf{u}_i + t_3(\mathsf{x})\right)\mathsf{v}_i = \mathsf{R}(\mathsf{x})\,\mathsf{u}_i + \mathsf{t}(\mathsf{x}), \tag{24}$$

where $\mathsf{x} \in \mathbb{R}^{12}$, r_i are the row vectors of $\mathsf{R}(\mathsf{x})$, and t_i are i^{th} elements of $\mathsf{t}(\mathsf{x})$. Notice again that (24) is analogous to (10), hence its residual error can be written as in (11). On the other hand, a convex relaxation of the constraint $\mathsf{R}(\mathsf{x}) \in SO(3)$ can be expressed as a LMI, using Theorem 5.2.

Corollary 6.2 $\mathsf{R}(\mathsf{x}) \in conv(SO(3))$ *is possible if and only if the following LMI is feasible for some* $\mathsf{x} \in \mathbb{R}^{12}$:

$$\mathcal{K}_a = \left\{\mathsf{I}_{4\times4} + \mathcal{L}(\mathsf{R}(\mathsf{x}))\right\} \succeq 0. \tag{25}$$

6.3. Relative Pose

We consider the homogeneous vector pairs $\{\mathsf{u}_i, \mathsf{v}_i\} \in \mathbb{R}^3$, $i = 1, \ldots, n$, with $n \geq 8$, which are the measurements of the image points $\{U_i, V_i\}$ of two calibrated cameras. For an essential matrix E, the relationship between two image points can be expressed as,

$$\left(\left[(\mathsf{n}_i)_1\mathsf{e}_2(\mathsf{x}) - (\mathsf{n}_i)_2\mathsf{e}_1(\mathsf{x})\right]^\intercal \mathsf{u}_i\right)\mathsf{v}_i = [\mathsf{n}_i]_\times \mathsf{E}(\mathsf{x})\,\mathsf{u}_i, \tag{26}$$

where $\mathsf{x} \in \mathbb{R}^9$, $[\mathsf{n}_i]_\times$ is a $3{\times}3$ skew symmetric matrix for any $\mathsf{n}_i \in \text{Null}(\mathsf{v}_i^\intercal)$, and $\mathsf{e}_\mathsf{i}(\mathsf{x})$ are the row vectors of E. Notice again that (26) is analogous to (10), hence its residual error can be written as in (11). On the other hand, a direct application of Proposition 5.8 provides a convex relaxation, as a LMI, to the Essential matrix constraint, i.e. $\mathsf{E}(\mathsf{x}) \in \mathcal{E}$.

Corollary 6.3 $\mathsf{E}(\mathsf{x}) \in \mathcal{E}$ *is possible only if the following LMI is feasible for some* $\mathsf{x} \in \mathbb{R}^9$:

$$\mathcal{K}_r = \left\{2\,\mathsf{I}_{4\times4} + \mathcal{L}(\mathsf{E}(\mathsf{x}))\right\} \succeq 0. \tag{27}$$

7. Experiments & Results

We performed experiments for the three problems described in Section 6, both on synthetic and real data. Our approach was implemented in MATLAB2016a using the Yalmip[1] toolbox and Mosek[2] as SDP solver. All experiments were carried out on an Intel Core i7 CPU 2.60GHz with 12GB RAM. Although there is still room for improvement by correctly modeled covariance matrix of individual applications, we used the Euclidean distance, i.e. $\Sigma = I$ in Section 4.1. The error measurement metrics used for evaluating the quality of the results are: the errors in rotation R, translation T, scale S, and the RMS 3D error. For each experiment, we compute the errors $\Delta\mathsf{r} = ||\mathsf{r} - \mathsf{r}_{gt}||$, $\Delta\mathsf{t} = ||\mathsf{t} - \mathsf{t}_{gt}||$, and $\Delta s = ||s - s_{gt}||$. The errors reported as $\Delta\,R$, $\Delta\,T$, and $\Delta\,S$ are the RMS values of N experiments. Here, r is a vector obtained by stacking three rotation angles in degrees, and r_{gt}, t_{gt} and s_{gt} are the ground truth values.

[1] https://yalmip.github.io/
[2] https://www.mosek.com/

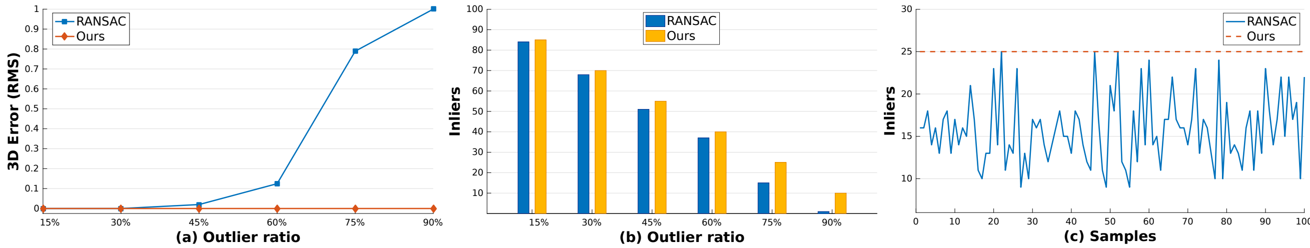

Figure 2: **Similarity Transformation** (synthetic data): RANSAC vs. Ours. Plots (a) and (b) show the behavior of different runs for both methods, while increasing the outlier ratio. Plot (c) shows multiple instances of RANSAC results against our method. RANSAC results in different number of inliers, while our method provides always the same exact solution.

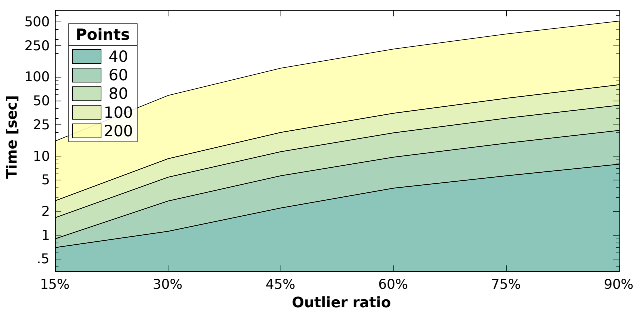

Figure 1: **Similarity Transformation** (synthetic data): Runtime of our method with increasing number of points and outlier ratio.

Outliers	N=30		N=40		N=50	
	W/	W/O	W/	W/O	W/	W/O
15 %	0.26	0.61	0.44	2	0.64	7
30 %	0.62	9	1.17	47	2.05	315
45 %	1.24	48	2.68	467	4.38	-
60 %	2.30	498	4.63	-	7.84	-
75 %	3.42	-	7.04	-	11.81	-

W/: with constraints (Ours). **W/O**: without constraints. **(−)**: greater than 3600 sec.

Table 2: **Similarity Transformation** (synthetic data): Time comparison between the solutions obtained with and without the LMI constraint (23). Even for small N (number of points), MI-SDP (8) method without LMI constraints takes a very long time.

7.1. Similarity Transformation

In this section, we show the general properties of the proposed method before and after adding the LMI constraints. For all our experiments, we restrict the reconstruction scale within $[0.2, 5.0]$, recall α in Proposition 5.5.

Synthetic data. We synthetically generated Similarity Transformations, which were applied to N points (uniformly generated) in order to obtain Ground Truth correspondences. We then introduced outliers by adding a high amount of noise to a particular subset of these correspondences, until the desired outlier ratio was obtained.

In Table 2, a time comparison between MI-SDP (without the LMI constraints from equation (8)) and Ours (with constraints) has been presented for a small number of points. In the presence of high amount of outliers, the speedups are very significant. Fig. 1 shows the runtime of our method with the constraints. In Fig. 2, we have compared our method against RANSAC. We fixed the number of points to $N = 100$ for all cases, and ran different instances while increasing the outlier ratio. For Fig. 2c the outlier ratio was set to 75% and the experiments were conducted for 100 times. Notice that the average performance of RANSAC (in this figure) corresponds approximately to its performance for 75% in Fig. 2b.

Real data. Images from the Yahoo Flickr Creative Commons dataset [18, 32] were used to obtain 3D Reconstructions with COLMAP[3], an open-source Structure-from-Motion (SfM) pipeline [30]. The SfM Reconstructions acquired correspond to: **Colosseum** (2060 images), **Notre Dame** (3743 images), **Pantheon** (1385 images) and **Trevi Fountain** (2909 images). Approximately only $10\,\%$ of the images had GPS information; the numbers of valid GPS Tag found for each dataset is shown in Table 3. This table also provides additional quantitative results. Due to the lack of Ground Truth registration parameters, the reported quantitative results were visually evaluated. As expected, the quality of the results improves with increasing numbers of inliers. This can be observed with a large number of inliers in Colosseum, in contrast to the Pantheon dataset (where the number of inliers is only 14). For the qualitative evaluation, all 3D Point Clouds were registered to Open Street Maps[4] and are shown in Fig. 3.

Fig. 4 provides a comparison between RANSAC and our method for the Colosseum dataset. In the same figure, the GPS elevation measurements plot is also provided. It can be observed that the GPS data exhibit huge deviations along the vertical axis, affecting RANSAC in particular. Conduct-

[3] https://colmap.github.io/
[4] http://openstreetmap.org

Scene	$\Delta\,\theta$ (Yaw)	$\Delta\,\theta$ (Pitch)	$\Delta\,\theta$ (Roll)	$\Delta\,T$	Height	Scale	$\lvert\zeta^*\rvert\,/\,N$	Time [sec]
Colosseum	$<\,1^o$	$<\,1^o$	$<\,1^o$	$<\,1\,m$	$<\,1\,m$	$<\,1\%$	$117/147$	$88.26\,s$
Notre Dame	$<\,3^o$	$<\,2^o$	$<\,1^o$	$<\,2\,m$	$<\,1\,m$	$<\,1\%$	$103/144$	$43.17\,s$
Pantheon	$<\,3^o$	$<\,5^o$	$<\,2^o$	$<\,3\,m$	$<\,2\,m$	$<\,7\%$	$14/\;\;47$	$16.12\,s$
Trevi Fontain	$<\,2^o$	$<\,1^o$	$<\,3^o$	$<\,1\,m$	$<\,1\,m$	$<\,3\%$	$104/140$	$65.68\,s$

$\Delta\,\theta$ [degree]: rotation error (Yaw, Pitch and Roll). $\Delta\,T$ [meters]: translation error. ζ^*: maximum consensus set. N: number of available GPS tags.

Table 3: **Similarity Transformation** (real data): Quantitative results. Visually evaluated quality of registration.

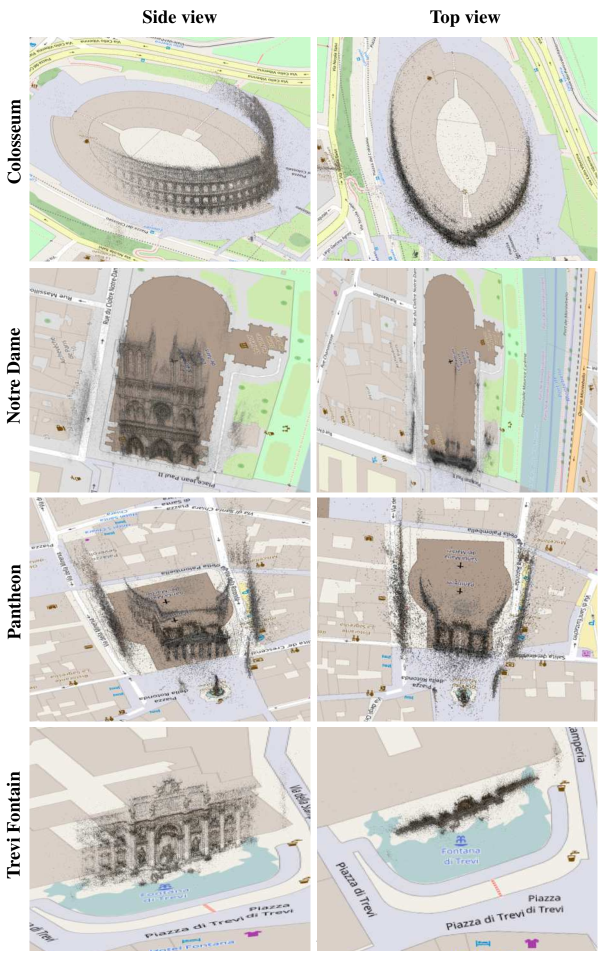

Figure 3: **Similarity Transformation** (real data): Side and Top views of the registered 3D Reconstructions obtained using our method with GPS information.

ing more experiments, the RANSAC results were often of very poor quality (sometimes even completely below the ground plane), while our method consistently provided the exact same solution.

7.2. Absolute Pose

For this experiment, we used a SfM Reconstruction from an unordered set of 30 images. The ground truth absolute pose [R|t] for 6 query images was also provided in the dataset. We follow an established [19] image-to-SfM localization pipeline to find the absolute pose of the query images: we compute SIFT [23] features in all query images, and match descriptors against the database of SIFT features from all 30 images used in the reconstruction. Since descriptors used in the SfM Reconstruction are associated with 3D points, we obtain a list of potential 3D-2D correspondences $\{\mathsf{u}_i, \mathsf{v}'_i\}$. The 2D image points v'_i are transformed into normalized (homogeneous) image coordinates: $\mathsf{v}_i = \mathsf{K}^{-1}\,[\mathsf{v}'_i\;1]^T$, where K denotes internal camera calibration matrix. The list of correspondences $\{\mathsf{u}_i, \mathsf{v}_i\}$ for each query image contains on average an outlier percentage of $44.25\,\%$. We aim at recovering the absolute pose [R|t] of the query images, given the list of potential matches, as described in Section 6.2.

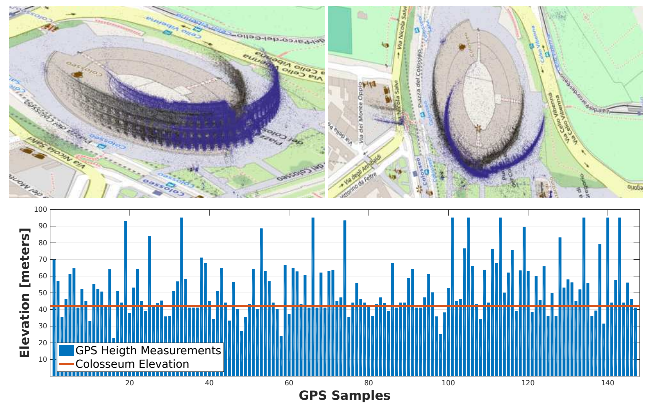

Figure 4: **Similarity Transformation** (real data): comparison between RANSAC (purple) vs. Ours. The bottom plot shows the GPS elevation measurements, varying from 20 to 90 meters, where 42 (straight line) is the true elevation of Colosseum.

Fig 5 shows a comparison with others methods, namely: ASPnP [39], DLT+LM[5] [17], EPnP [20], REPPnP [12]. The low rotational and transitional errors reported are a direct consequence of the inliers-set consensus maximization. Table 4 complements this information with the execution time and the numbers of inliers. Here, we provide

[5]Direct Linear Transform (DLT) followed by Levenberg-Marquardt (LM) minimization of the reprojection error.

Scene	Image	Method	$\lvert\zeta\rvert$ / N	ΔR [degree]	ΔT [%]	Time [sec]
Fountain		RANSAC	20 / 39	0.29	4.81	**0.61**
		Ours	**25** / 39	**0.15**	**1.76**	3.35
Herz-Jesu		RANSAC	35 / 70	2.12	3.20	**0.63**
		Ours	**49** / 70	**0.12**	**2.87**	23.84

$\lvert\zeta\rvert$: number of inliers. **Ours**: method with constraints. ΔR [degree]: rotation error. ΔT [%]: translation error.

Table 5: **Relative Pose** (real data): RANSAC vs. Our method with and without LMI constraints.

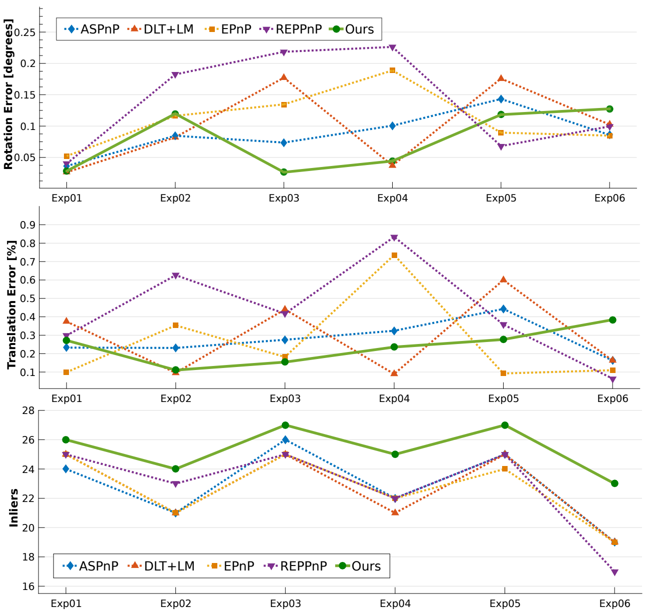

Figure 5: **Absolute Pose** (real data): comparison with other methods. Rotation Error, Translation Error and Number of True Inliers are shown for Six Experiments.

	Error		Inliers	Time [sec]	
	ΔR	ΔT	$\lvert\zeta^*\rvert$ / N	$Ours$	$Ours^*$
Exp01	0.03	0.27	26 / 42	13.86	2.96
Exp02	0.12	0.28	24 / 45	25.71	2.78
Exp03	0.12	0.11	27 / 46	50.27	13.80
Exp04	0.13	0.41	25 / 46	61.18	20.44
Exp05	0.04	0.24	27 / 47	174.81	10.40
Exp06	0.18	0.31	23 / 44	120.24	58.06

ΔR [degree]: rotation error. ΔT [%]: translation error. ζ^*: maximum consensus. $Ours^*$: imposing the additional constraint (16): $R + R^T \succeq I$.

Table 4: **Absolute Pose** (real data): Error, Inliers Set and Time for our method. By adding an additional constraint, we obtained *2x* speedup in **Exp06** (worst case) and *16x* in **Exp05** (best case).

two sets of results, one with only the LMI constraint (25) and the other with an extra restriction on rotation angle (of $|\theta| \leq 60°$). This table confirms once again the key contribution of this paper: with more restrictions imposed using LMIs, the BnB explores less space – reducing significantly the time in our experiments – while finding the same solution. Note that the rotation angle restriction used in these experiments may not always be true for Absolute Pose problem. Nevertheless, with only some vague prior knowledge (such as IMU measurements) this restriction can be confirmed.

7.3. Relative Pose

We conducted experiments with two different datasets – Fountain and Herz-Jesu – form [31]. The details of the datasets and the obtained results are shown in Table 5. The reported results obtained by our method in presence of the LMI constraint (27) is compared against RANSAC conducted on 5-point algorithm [25]. One can observe that the global search method with LMI constraints finds a larger set of inliers than that of RANSAC, for both datasets. The quality of relative pose estimation improves with increasing inliers in both cases.

8. Conclusion

We proposed a general global optimization framework for consensus maximization with linear matrix inequality constraints. We derived several LMI constraints and demonstrated that a number of central computer vision problems can be cast into this form. In particular, we successfully conducted experiments on problems of similarity transformation, absolute pose, and relative pose estimation. Experiments demonstrated also a significant speedup in computation time due to the addition of the LMI constraints, under a globally optimal framework.

As future work it is worth to explore the use of LMI constraints in combination with other exact methods [5, 10, 22, 26, 38], since the LMI constraints have the potential to improve their effectiveness. We therefore see these methods as complementary work, rather than competitors. Apart from combining it with other optimization methods, we further look forward to explore other computer vision problems that fit into the proposed optimization framework.

References

[1] J. C. Bazin, Y. Seo, R. I. Hartley, and M. Pollefeys. Globally optimal inlier set maximization with unknown rotation and focal length. In *ECCV*, pages 803–817, 2014.

[2] J. C. Bazin, Y. Seo, and M. Pollefeys. Globally optimal consensus set maximization through rotation search. In *ACCV*, pages 539–551, 2012.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[4] T.-J. Chin, Y. Heng Kee, A. Eriksson, and F. Neumann. Guaranteed outlier removal with mixed integer linear programs. In *CVPR*, June 2016.

[5] T.-J. Chin, P. Purkait, A. Eriksson, and D. Suter. Efficient globally optimal consensus maximisation with tree search. In *CVPR*, pages 2413–2421, 2015.

[6] J. W. Chinneck. *Feasibility and Infeasibility in Optimization: Algorithms and Computational Methods*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[7] S. Choi, T. Kim, and W. Yu. Performance evaluation of ransac family. 2009.

[8] O. Chum and J. Matas. Matching with prosac - progressive sample consensus. In *CVPR*, pages 220–226, 2005.

[9] M. Conforti, G. Cornuejols, and G. Zambelli. *Integer Programming*. Springer Publishing Company, Incorporated, 2014.

[10] O. Enqvist, E. Ask, F. Kahl, and K. Åström. *Robust Fitting for Multiple View Geometry*, pages 738–751. 2012.

[11] D. Eppstein. Quasiconvex programming. *Combinatorial and Computational Geometry*, 52:287–331, 2005.

[12] L. Ferraz, X. Binefa, and F. Moreno-Noguer. Very fast solution to the pnp problem with algebraic outlier rejection. In *CVPR*, 2014.

[13] M. A. Fischler and R. 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.

[14] J. Fredriksson, V. Larsson, and C. Olsson. Practical robust two-view translation estimation. In *CVPR*, pages 2684–2690, 2015.

[15] A. Habed, D. Pani Paudel, C. Demonceaux, and D. Fofi. Efficient pruning lmi conditions for branch-and-prune rank and chirality-constrained estimation of the dual absolute quadric. *CVPR*, pages 493–500, 2014.

[16] R. I. Hartley and F. Kahl. Global optimization through rotation space search. *IJCV*, 82(1):64–79, 2009.

[17] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[18] J. Heinly, J. L. Schönberger, E. Dunn, and J.-M. Frahm. Reconstructing the World* in Six Days *(As Captured by the Yahoo 100 Million Image Dataset). In *CVPR*, 2015.

[19] A. Irschara, C. Zach, J. M. Frahm, and H. Bischof. From structure-from-motion point clouds to fast location recognition. In *CVPR*, pages 2599–2606, June 2009.

[20] V. Lepetit, F.Moreno-Noguer, and P.Fua. Epnp: An accurate o(n) solution to the pnp problem. *International Journal Computer Vision*, 81(2), 2009.

[21] H. Li. A practical algorithm for L triangulation with outliers. In *CVPR*, 2007.

[22] H. Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *ICCV*, pages 1074–1080, Sept 2009.

[23] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[24] J. Matoušek, M. Sharir, and E. Welzl. A subexponential bound for linear programming. *Algorithmica*, 16(4-5):498–516, 1996.

[25] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–777, June 2004.

[26] C. Olsson, O. Enqvist, and F. Kahl. A polynomial-time bound for matching and registration with outliers. In *CVPR*, pages 1–8, 2008.

[27] D. P. Paudel, A. Habed, C. Demonceaux, and P. Vasseur. Robust and optimal sum-of-squares-based point-to-plane registration of image sets and structured scenes. In *ICCV*, pages 2048–2056, 2015.

[28] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J.-M. Frahm. Usac: A universal framework for random sample consensus. *TPAMI*, 35(8):2022–2038, Aug. 2013.

[29] R. Sanyal, F. Sottile, and B. Sturmfels. Orbitopes. *Mathematika*, 57(02):275–314, 2011.

[30] J. L. Schönberger and J.-M. Frahm. Structure-from-motion revisited. In *CVPR*, 2016.

[31] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, pages 1–8, June 2008.

[32] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. Yfcc100m: The new data in multimedia research. *Commun. ACM*, 59(2):64–73, Jan. 2016.

[33] B. Tordoff and D. W. Murray. Guided sampling and consensus for motion estimation. In *ECCV*, pages 82–98, London, UK, UK, 2002. Springer-Verlag.

[34] P. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138 – 156, 2000.

[35] C. Vinzant. What is... a spectrahedron. *Notices Amer. Math. Soc*, 61(5):492–494, 2014.

[36] L. A. Wolsey. *Integer programming*. Wiley-Interscience series in discrete mathematics and optimization. J. Wiley & sons, New York (N.Y.), Chichester, Weinheim, 1998. A Wiley-Interscience publication.

[37] J. Yang, H. Li, and Y. Jia. Optimal essential matrix estimation via inlier-set maximization. In *ECCV*, pages 111–126, 2014.

[38] Y. Zheng, S. Sugimoto, and M. Okutomi. Deterministically maximizing feasible subsystem for robust model fitting with unit norm constraint. In *CVPR*, pages 1825–1832, June 2011.

[39] Y. Zheng, S. Sugimoto, and M. Okutomi. Aspnp: An accurate and scalable solution to the perspective-n-point problem. *IEICE Transactions*, (7):1525–1535, 2013.