Consensus Maximization for Semantic Region Correspondences

Pablo Speciale[1], Danda P. Paudel[2], Martin R. Oswald[1],
Hayko Riemenschneider[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, hayko, vangool}@vision.ee.ethz.ch

Day / Night Registration | Outdoor / Indoor Registration | Scan / CAD Registration

Figure 1: Example registration results. Our approach solves challenging registration problems by maximizing the number of corresponding semantic regions – such as windows, doors or balconies – for datasets from different modalities, with large amounts of noise and outliers, little data overlap, or significantly different data statistics.

Abstract

We propose a novel method for the geometric registration of semantically labeled regions. We approximate semantic regions by ellipsoids, and leverage their convexity to formulate the correspondence search effectively as a constrained optimization problem that maximizes the number of matched regions, and which we solve globally optimal in a Branch-and-Bound fashion. To this end, we derive suitable linear matrix inequality constraints which describe ellipsoid-to-ellipsoid assignment conditions. Our approach is robust to large percentages of outliers and thus applicable to difficult correspondence search problems. In multiple experiments we demonstrate the flexibility and robustness of our approach on a number of challenging vision problems.

1. Introduction

Correspondence search is a fundamental subproblem of many computer vision tasks including pixel matching in 3D reconstruction as well as feature matching in shape matching, localization, retrieval or registration tasks. With the wide availability of reliable semantic classification algorithms for images or 3D data, our goal is to leverage semantic information to resolve ambiguities and improve the efficiency of correspondence search.

We target challenging correspondence problems with small data overlap, large amounts of noise and outliers, or the registration of datasets with significantly different data statistics. Figure 1 shows examples of such difficult registration problems. For instance, the registration of two 3D models captured at day and night light conditions is extremely challenging with only classical local features such as structure-from-motion (SfM) feature points [2, 46]. Another example is the registration of a building scan with a corresponding computer-aided design (CAD) model in which the data statistics of the two input meshes differ substantially. Nevertheless, higher semantic features such as windows, doors or balconies are nowadays easy to detect and more descriptive and robust than classical SfM features.

In this paper, we aim to unify such correspondence problems and consider the geometric registration of multiple compact regions with semantic labels that can be linked via an affine or projective transformation. We seek to estimate a transformation which maximizes the number of matching regions with the same semantic label. To account for a considerable amount of data variations, we approximate the semantically labeled regions with ellipsoids whose properties are also beneficial for an effective global optimization strategy. We derive necessary conditions for ellipsoid-to-ellipsoid inclusion test that can be embedded as constraints into consensus maximization problem.

Contributions. We propose a global optimization framework for semantic region assignments. To account for noise and outliers we approximate the semantic regions by ellipsoids and derive suitable linear matrix inequality (LMI) constraints that allow for ellipsoid-to-ellipsoid correspondence testing within a consensus maximization framework.

Acknowledgements. This work has received funding from the European Unions Horizon 2020 research and innovation programme project Built2Spec, and the European Research Council project VarCity, under grant agreements No. 637221 and 273940, respectively.

Due to the global optimization approach, our method allows for large amount of outliers. We demonstrate the versatility of our method in multiple experiments on real and synthetic data and show competitive registration results for two computer vision problems, namely: similarity transformation, and purely rotating cameras.

2. Related Work

Our work builds upon a large body of previous theoretical results for effective *consensus maximization*. Moreover, for the applications we consider, there are a handful of related works using *specialized registration methods* which are usually tailored for a particular problem case only.

Consensus Maximization. Distinguishing between model inliers and outliers and the maximization of the number of inliers has been a central computer vision problem from early on. Due to their effectiveness and low runtime, stochastic or **local methods** like RANSAC [20] and its variants [16, 35] gained great popularity. Although being effective for many tasks these methods have no optimality guarantees and are slow or break down entirely for large amounts of outliers that we consider in this work.

We therefore build on **global methods** for consensus maximization. The vast majority of global methods prunes the search paths during exhaustive search using the Branch-and-Bound (BnB) strategy, e.g. [4, 5, 10, 25, 29, 38, 45]. To speed up the BnB optimization, several methods combine it with Mixed Integer Programming (MIP) [13, 18, 29, 43]. As an alternative search strategy, Chin *et al*. [14] use A^*-search to traverse the solution space.

Specialized Registration Methods. Many works consider affine transformation problems or more specialized transformations during consensus maximization. For instance, rotations [5, 25], rotation+focal length [4], translation [21], rotation+translation [10], rotation+translation+scale [32] or essential matrices [44]. Speciale *et al*. [38] provide a more general framework handling most of these transformation types as long as they can be expressed by linear matrix inequality constraints. Closely related to our work, Paudel *et al*. [33] consider polygon to ellipsoid inclusions with different semantics in order to solve for 2D homographies or 3D projective transformations. In contrast to our approach they need some known semantic correspondences and do not use consensus maximization. Apart from some works which estimate correspondences among semantically similar objects or regions in different images [9, 23, 24, 40], there are few works which consider semantic information for registration problems.

Application-wise, there are a few methods which also targeted difficult registration problems, like **outdoor/indoor registration** of building scans [17]. The problem of **day/night registration** has been considered for aligning structure-from-motion models [34], image-based localization [46], image matching [3, 26, 47] and video registration [2].

In sum, there are mostly specially tailored solutions for solving particular registration problems and few of them are able to incorporate semantic information. Currently, there exists no generic method which tackles a larger class of such registration problems. Therefore, we aim to introduce a single generic approach which leverages semantic information and which handles a wide-range of applications.

3. Background and Notations

We denote matrices with upper case letters and their elements by double-indexed lower case letters: $\mathsf{A} = (a_{ij})$. The row-wise representation of $m \times n$ matrices are denoted by $\mathsf{A} = [\mathsf{a}_1, \dots, \mathsf{a}_i, \dots, \mathsf{a}_m]^\intercal$, where a_i are n-dimensional vectors. We express positive semi-definiteness (resp. positive-definiteness) of a symmetric matrix by $\mathsf{A} \succeq 0$ (resp. $\mathsf{A} \succ 0$). We further define a n-dimensional vector $\mathsf{e} = (0, \dots, 0, 1)^\intercal$ and the upper-left $(m-1) \times (n-1)$ block of A by $\hat{\mathsf{A}}$.

A key ingredient of our work is the so-called S-Procedure, which defines conditions under which a particular quadratic inequality is a consequence of another quadratic inequality.

Lemma 3.1 (S-Procedure [42]) *Let A_0 and A_1 be symmetric matrices. $\mathsf{x}^\intercal \mathsf{A}_0 \mathsf{x} \leq 0$ holds for all x which satisfy $\mathsf{x}^\intercal \mathsf{A}_1 \mathsf{x} \leq 0$, if there exists $\lambda \geq 0$ such that $\lambda \mathsf{A}_1 \succeq \mathsf{A}_0$.*

An important tool for converting some nonlinear matrix inequalities into linear inequalities is called Schur complement.

Lemma 3.2 (Schur Complement [27]) *A symmetric block-partitioned matrix $\mathsf{D} = \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\intercal & \mathsf{C} \end{bmatrix} \succeq 0$, if and only if both $\mathsf{A} \succeq 0$ and $\mathsf{C} - \mathsf{B}^\intercal \mathsf{A}^{-1} \mathsf{B} \succeq 0$.*

3.1. Region Inside an Ellipsoid

We represent source and target semantic regions with the help of ellipsoids. Given two ellipsoids, one from the under-approximation of the source region and other from the over-approximation of the target region, we are interest to know whether the source ellipsoid can be transformed such that it fits inside the target ellipsoid.

Definition 3.3 (Ellipsoid) *An ellipsoid $\mathcal{E}$ in a $(d-1)$-dimensional space can be represented by a $d \times d$ matrix $\mathsf{Q} \succeq 0$ whose $(d-1) \times (d-1)$ upper-left block $\hat{\mathsf{Q}}$ satisfies $\hat{\mathsf{Q}} \succ 0$. Using homogeneous coordinate vectors, in which points in $(d-1)$-space are represented by $\mathsf{x} \in \mathbb{R}^d$, $\mathcal{E}$ is defined by $\mathcal{E} = \{\mathsf{x} : \mathsf{x}^\intercal (\mathsf{Q} - \mathsf{e}\mathsf{e}^\intercal) \mathsf{x} \leq 0\}$.*

In this work, the outer and inner ellipsoids are estimated in the form of extremal volume ellipsoids: the minimum

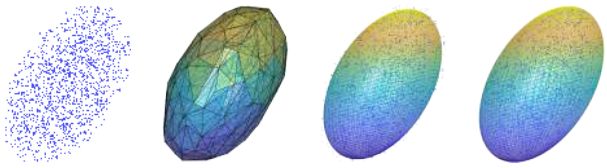

Figure 2: Definitions of Inner and Outer Ellipsoids. The left two images show a point cloud and its convex hull. The *inner ellipsoid* is the largest ellipsoid that fits into the convex hull. The *outer ellipsoid* is the smallest ellipsoid which encloses the point cloud (see Def. 3.4 and 3.5).

volume ellipsoid that covers a given set of points or the maximum volume ellipsoid that lies inside a convex polyhedron. An illustration of the approximated ellipsoids are shown in Fig 2.

Definition 3.4 (Outer Ellpsoid [7]) *The minimum volume ellipsoid – so called Löwner-John ellipsoid – of a compact and non-empty set $\mathcal{S} \subseteq \mathbb{R}^d$ is the outer ellipsoid $\overline{\mathcal{E}}$ that covers $\mathcal{S}$.*

For a convex set $\mathcal{S}$, the volume of an ellipsoid $\mathcal{E}$ being proportional to $\sqrt{\det(\hat{\mathsf{Q}}^{-1})}$ [8, p.48], the minimum volume ellipsoid can be obtained by solving the following concave maximization problem:

$$\begin{aligned} \max_{\mathsf{Q}} \quad & \log\det \hat{\mathsf{Q}} \\ \text{s.t.} \quad & \mathsf{x}^\intercal \left(\mathsf{Q} - \mathsf{e}\mathsf{e}^\intercal\right)\mathsf{x} \leq 0, \;\; \forall \mathsf{x} \in \mathcal{S}, \; \hat{\mathsf{Q}} \succ 0, \;\; \mathsf{Q} \succeq 0. \end{aligned} \tag{1}$$

Definition 3.5 (Inner Ellipsoid [7]) *The maximum volume ellipsoid for a non-empty polyhedron $\mathcal{S} = \{\mathsf{x} : \mathsf{a}_i^\intercal \mathsf{x} \leq 0, i = 1, \dots n\}$ is the inner ellipsoid $\underline{\mathcal{E}}$ enclosed within $\mathcal{S}$.*

As in (1), the optimal solution for the maximum volume ellipsoid can be obtained using interior point methods [7], by solving the following concave minimization problem:

$$\begin{aligned} \min_{\mathsf{Q}} \quad & \log\det \hat{\mathsf{Q}} \\ \text{s.t.} \quad & \mathsf{a}_i^\intercal \left(\mathsf{Q} - \mathsf{e}\mathsf{e}^\intercal\right)\mathsf{a}_i \leq 0, \quad \forall\, i, \; \hat{\mathsf{Q}} \succ 0, \;\; \mathsf{Q} \succeq 0. \end{aligned} \tag{2}$$

3.2. Ellipsoid under Projective Transformation

Consider a linearly parameterized projective transformation matrix $\mathsf{H} \in \mathbb{R}^{d\times d}$. This transformation relates source and target data points, with homogeneous coordinate vectors y and x respectively, by $\mathsf{y} \simeq \mathsf{H}\mathsf{x}$. Typically, the problem of source to target data registration is to estimate unknown H from known correspondences between points. In our work, such correspondences are unknown. Instead, we consider the correspondences are given in the form of semantic regions, which are approximated with the help of ellipsoids. Now, we are interested to establish the relationship between ellipsoids under the transformation H.

When an ellipsoid $\mathcal{E} = \{\mathsf{x} : \mathsf{x}^\intercal(\mathsf{Q} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{x} \leq 0\}$ undergoes a projective transformation $\mathsf{y} \simeq \mathsf{H}\mathsf{x}$, the transformed ellipsoid can be expressed as,

$$\mathcal{E}(\mathsf{H}) = \{\mathsf{y} : \mathsf{y}^\intercal \mathsf{H}^{-\intercal}(\mathsf{Q} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{H}^{-1}\mathsf{y} \leq 0\}, \tag{3}$$

such that $(\mathsf{H}\mathsf{x})^\intercal \mathsf{H}^{-\intercal}(\mathsf{Q} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{H}^{-1}(\mathsf{H}\mathsf{x}) \leq 0$ is satisfied.

4. The Consensus Maximization Problem

Given the putative correspondences between two sets of semantic regions, one from the source and the other from target data, we aim to maximize their consensus such that their exists a geometric transformation matrix H. Recall that we represent the source and target regions by inner and outer ellipsoids, respectively. Let an unknown projective transformation matrix H that relates a known pair of such ellipsoids $\mathcal{P} = \{\underline{\mathcal{E}}, \overline{\mathcal{E}}\}$. We refer $\mathcal{P}$ as a putative assignment, if it is not guaranteed to be true, rather it is a candidate that needs to be probed for its validity. Then, the problem of consensus maximization for semantic region correspondences can be stated as follows:

Problem 4.1 *Given a known set $\mathcal{S} = \{\mathcal{P}_i\}_{i=1}^n$,*

$$\begin{aligned} \max_{\mathsf{H}, \zeta \subseteq \mathcal{S}} \quad & |\zeta|, \\ \text{s.t.} \quad & \underline{\mathcal{E}}_i \subseteq \overline{\mathcal{E}}_i(\mathsf{H}), \quad \forall \mathcal{P}_i \in \zeta. \end{aligned} \tag{4}$$

This problem, however, is difficult to solve due to its combinatorial nature. Our following feasibility conditions for ellipsoid-to-ellipsoid correspondences are important tools for solving this problem in an efficient manner.

4.1. Ellipsoid-to-Ellipsoid Assignment Conditions

Proposition 4.2 *Let $\mathcal{P} = \{\underline{\mathcal{E}}, \overline{\mathcal{E}}\}$ be a pair of corresponding ellipsoids, defined as $\underline{\mathcal{E}} = \{\mathsf{y} : \mathsf{y}^\intercal(\underline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{y} \leq 0\}$ and $\overline{\mathcal{E}} = \{\mathsf{x} : \mathsf{x}^\intercal(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{x} \leq 0\}$. The ellipsoids are related by a projective transformation $\mathsf{y} \simeq \mathsf{H}\mathsf{x}$ by $\underline{\mathcal{E}} \subseteq \overline{\mathcal{E}}(\mathsf{H})$ if and only if, there exists a scalar $\lambda \geq 0$ such that,*

$$\lambda(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal) \succeq \mathsf{H}^\intercal(\underline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{H}. \tag{5}$$

Proof $\underline{\mathcal{E}} \subseteq \overline{\mathcal{E}}(\mathsf{H}) \implies \mathsf{y}^\intercal(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{y} \leq 0$ for every $\mathsf{y} : \mathsf{y}^\intercal \mathsf{H}^{-\intercal}(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{H}^{-1}\mathsf{y} \leq 0$. Now, from Lemma 3.1, $\underline{\mathcal{E}} \subseteq \overline{\mathcal{E}}(\mathsf{H})$ iff $\exists \lambda \geq 0 : \lambda \mathsf{H}^{-\intercal}(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal)\mathsf{H}^{-1} \succeq \underline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal$. This condition turns to (5) under the similarity transformation[1] with a full rank matrix H. ∎

In general, the *feasibility test* of (5) for known $\mathcal{P}$ and unknown H is a non-convex problem. However, it is still possible to derive its convex relaxations for some specific problems. Please, refer to the supplementary materials for few such relaxation examples.

[1] In linear algebra, $P^{-1}\,A\,P$ is a *similarity transformation* of matrix A.

In this paper, we focus on the cases when the matrix H represents affine transformations. Note that for affine matrices, the last row of H takes the form $\mathsf{h}_d = \mathsf{e}$. More importantly, the assignment condition of (5), under affine transformations, can be expressed as a *Linear Matrix Inequality* (LMI)[2]. The feasibility of such LMIs can be tested using *Semi-Definite Programming* (SDP). Our following proposition offers the ellipsoid-to-ellipsoid assignment conditions in the form of LMIs.

Proposition 4.3 *For the Cholesky decomposition of positive semi-definite matrix $\underline{\mathsf{Q}} = \underline{\mathsf{L}}^\mathsf{T}\underline{\mathsf{L}}$ (inner ellipsoid) and an affine matrix H, the following statements are equivalent:*

$$\begin{aligned} &(i) \qquad \exists\lambda: \quad \lambda(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\mathsf{T}) \succeq \mathsf{H}^\mathsf{T}(\underline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\mathsf{T})\mathsf{H}. \\ &(ii) \qquad \exists\lambda: \quad \begin{bmatrix} \mathsf{I}_{d\times d} & \underline{\mathsf{L}}\mathsf{H} \\ (\underline{\mathsf{L}}\mathsf{H})^\mathsf{T} & \lambda(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\mathsf{T}) + \mathsf{e}\mathsf{e}^\mathsf{T} \end{bmatrix} \succeq 0. \end{aligned} \tag{6}$$

Proof One can obtain the equivalence directly by applying Lemma 3.2 on statement (ii) for $\mathsf{h}_d = \mathsf{e}$. ∎

4.2. Mixed-Integer Programming

Using the proposed LMI conditions for ellipsoid-to-ellipsoid assignments, the Problem 4.1 for semantic consensus maximization can be expressed as a Mixed Integer Semi-Definite Program (MI-SDP) [18, 43], as in [38]. In this regard, we represent inlier/outlier assignments as binary variables for each putative correspondences, whereas the assignment conditions are expressed as LMI constraints. The MI-SDP then jointly searches for the binary variables as well as the transformation matrix such that the maximum number of assignment conditions are satisfied. We present this idea more precisely in our following preliminary result.

Result 4.4 *The Problem 4.1 can be solved optimally for the affine matrix H, binary variables z_i, scalar λ, and a sufficiently large positive semi-definite matrix $\mathbf{M}$, by solving the following MI-SDP,*

$$\begin{aligned} \min_{\mathsf{H},z_i,\lambda} \quad & \sum_{i=1}^{n} z_i, \\ \text{s.t.} \quad & \begin{bmatrix} \mathsf{I}_{d\times d} & \underline{\mathsf{L}}_i\mathsf{H} \\ (\underline{\mathsf{L}}_i\mathsf{H})^\mathsf{T} & \lambda(\overline{\mathsf{Q}}_i - \mathsf{e}\mathsf{e}^\mathsf{T}) + \mathsf{e}\mathsf{e}^\mathsf{T} \end{bmatrix} \succeq -z_i\,\mathbf{M}, \\ & \lambda \geq 0, z_i \in \{0,1\} \qquad \forall i. \end{aligned} \tag{7}$$

Note that a common λ is sufficient for all the assignments. This is because, if Eq. (5) is true for any assignment with some λ_i, it must also be true for any $\lambda \geq \lambda_i$. Here, we seek a single λ such that $\lambda \geq \lambda_i, \forall i$. Although (7) is still a large combinatorial problem, the optimal search of its variables can be performed efficiently using a Branch-and-Bound (BnB) paradigm specifically designed for MI-SDP.

One can observe from Eq. (7) that the putative assignment $\mathcal{P}_i$ is an inlier if $z_i = 0$, and an outlier otherwise. It is important to notice the problem of (7) is always feasible when $z_i = 1, \forall i$, irrespective to the legitimacy of the assignments. Similarly, all the assignment conditions must be satisfied if (7) is feasible with $z_i = 0, \forall i$. Therefore, we maximize the inlier set by minimizing the sum of z_i for all the assignments. In this case, the sufficiently large positive semi-definite matrix $\mathbf{M}$ helps us to ignore the constraints that arise form outlier measurements. In fact, it is a common practice in optimization to ignore invalid constraints by using a constant such as $\mathbf{M}$. See [15, Ch. 7] for guidelines on selecting this constant.

4.3. Multiple Regions with Same Semantics

In practice, both source and target data consist of multiple regions with same semantics. In such cases, it is difficult to establish one-to-one putative correspondences between region, using only the knowledge about semantic labels. Therefore, we assign every region from the source data to all the target regions with the same semantics. However, we are interested to only those solutions which also respect the one-to-one assignment criteria.

Let $l_j, j = 1, \ldots, s$ be the semantic labels in the source data and $\mathcal{L}(\mathcal{E})$ be the label of the ellipsoid $\mathcal{E}$. For every label l_j, we define a set of ellipsoids in the source data by $\mathcal{S}_j = \{\underline{\mathcal{E}} : \mathcal{L}(\underline{\mathcal{E}}) = l_j\}$ and in the target data by $\mathcal{T}_j = \{\overline{\mathcal{E}} : \mathcal{L}(\overline{\mathcal{E}}) = l_j\}$. Then, the assignments for label l_j is given by a set $\mathcal{A}_j = \mathcal{S}_j \times \mathcal{T}_j$, where $\times$ refers to the Cartesian product. The set of all putative assignments is given by $\mathcal{P} = \bigcup_{j=1}^{s} A_j$, where every pair $\mathcal{P}_i \in \mathcal{P}$ is a candidate assignment.

We now state our main result:

Result 4.5 *Assume that we are given semantic labels $l_j, j = 1, \ldots, s$ and a set of putative assignments $\mathcal{P}$, whose inlier assignments must respect an affine transformation $\mathsf{H} \in \mathbb{R}^{4\times4}$, with $\mathsf{h}_4 = \mathsf{e}$. For a binary decision variable z_i corresponding to every pair $\mathcal{P}_i \in \mathcal{P}$, an unknown scalar λ, and a sufficiently large known positive semi-definite matrix $\mathbf{M}$, the consensus among all the pairs $\mathcal{P}_i \in \mathcal{P}$ can be obtained by solving the following MI-SDP,*

$$\begin{aligned} \min_{\mathsf{H},z_i,\lambda} \quad & \sum_{\mathcal{P}_i\in\mathcal{P}} z_i, \\ \text{s.t.} \quad & \begin{bmatrix} \mathsf{I}_{4\times4} & \underline{\mathsf{L}}_i\mathsf{H} \\ (\underline{\mathsf{L}}_i\mathsf{H})^\mathsf{T} & \lambda(\overline{\mathsf{Q}}_i - \mathsf{e}\mathsf{e}^\mathsf{T}) + \mathsf{e}\mathsf{e}^\mathsf{T} \end{bmatrix} \succeq -z_i\,\mathbf{M}, \\ & \lambda \geq 0, z_i \in \{0,1\} \qquad \forall i, \\ & \sum_{\mathcal{P}_i\in\mathcal{A}_j(\underline{\mathcal{E}})} (1 - z_i) \leq 1, \qquad \forall\underline{\mathcal{E}} \in \mathcal{S}_j, \forall j, \\ & \sum_{\mathcal{P}_i\in\mathcal{A}_j(\overline{\mathcal{E}})} (1 - z_i) \leq 1, \qquad \forall\overline{\mathcal{E}} \in \mathcal{T}_j, \forall j, \end{aligned} \tag{8}$$

[2] *Linear Matrix Inequality* is a constraint on y such that $\mathsf{A}(\mathsf{y}) \succeq 0$, where $\mathsf{A}(\mathsf{y}) = \mathsf{A}_0 + \sum_i^n y_i\mathsf{A}_i$ for $\mathsf{A}_i \succeq 0\ \forall\, i$, and $\mathsf{y} = [y_1, \ldots, y_n]^\mathsf{T} \in \mathbb{R}^n$.

where, $\mathcal{A}(\mathcal{E})$ are all the assignments involving ellipsoid $\mathcal{E}$.

Note that the task of enforcing one-to-one assignment in (8) is addressed by following these two simple rules: (i) every ellipsoid from source data must have no more than one valid assignment; (ii) every ellipsoid from target data must have no more than one valid assignment. Recall, if the binary variable $z_i = 0$, the assignment $\mathcal{P}_i$ is an inlier. Otherwise, $\mathcal{P}_i$ is an outlier. In practice, enforcing one-to-one assignment criteria not only generates geometrically meaningful results, but also speeds up the MI-SDP significantly.

5. Applications

In this section, we specialize our Result 4.5 to the problems of similarity transformation and pure rotation estimation with additional problem-specific constraints.

5.1. SfM Reconstruction to Euclidean Scene

Let us consider that the inner and outer ellipsoids, $\underline{\mathcal{E}}_i$ and $\overline{\mathcal{E}}_i$, are extracted form the Structure-from-Motion (SfM) reconstruction and its Euclidean counterpart, respectively. Given assignments $\mathcal{P}_i$ based on their semantic labels, we wish to estimate the transformation matrix H that maximizes the assignment' consensus. In this particular case, $\mathsf{H} \in \mathbb{R}^{4\times4}$ is a similarity matrix, therefore offers an additional constraint that can be expressed as an LMI [38]. Note that similarity transformation is represented by a scaled-rotation matrix and a translation vector. The following definition deals with the structure of a scaled-rotation matrix.

Definition 5.1 (SSO(3)) *Given a real, compact, linear algebraic group $\mathcal{Q}$, a 3-dimensional scaled-special orthogonal group is defined by,*

$$SSO(3) = \{\mathsf{Q} \in \mathcal{Q} : \mathsf{Q}\mathsf{Q}^\intercal = \alpha^2\, \mathsf{I}_{3\times3},\ det(\mathsf{Q}) = \alpha^3, \alpha > 0\}, \tag{9}$$

Recall that the upper-left block $\hat{\mathsf{H}}$ of H, must satisfy the Definition 5.1. Now, the following theorem provides us a convex relaxation for $\hat{\mathsf{H}}$ as an LMI.

Theorem 5.2 (SSO(3) Orbitope [38]) *A 3×3 matrix $\mathsf{Q} \in SSO(3)$, only if there exists a scalar $\alpha > 0$:*

$$\alpha\, \mathsf{I}_{4\times4} + \mathcal{L}(\mathsf{Q}) \succeq 0, \tag{10}$$

for a linear function $\mathcal{L} : \mathbb{R}^{3\times3} \to \mathbb{R}^{4\times4}$ is defined by,

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

For a sufficiently large α, Eq. (10) is always satisfied. However, α is not an arbitrary scalar value, but the scale of the reconstruction. Given a rough knowledge about the scale of the reconstruction, α can be bounded. In practice, such bounds can be computed only form some vague prior knowledge, such as IMU/GPS measurements, or even from the extracted semantics. Once α is bounded, (10) turns out to be very useful during MI-SDP. We solve the problem of similarity transformation estimation using our Result 4.5 with additional constraint $\alpha\, \mathsf{I}_{4\times4} + \mathcal{L}(\mathsf{Q}) \succeq 0$ and $\underline{\alpha} \leq \alpha \leq \overline{\alpha}$. Where, $\underline{\alpha}$ and $\overline{\alpha}$ are known lower and upper bounds of the reconstruction scale, respectively.

5.2. Purely Rotating Cameras

The problem of pure rotation estimation appears while dealing with cases such as pan-tilt-zoom (PTZ) cameras [30] or image stitching [39] for panoramas. In this context, we assume that the cameras are calibrated and their measurements are given in the camera coordinate frame. Let $\mathcal{S} = \{\hat{\mathsf{u}}_i\}_{i=1}^n$ and $\mathcal{T} = \{\hat{\mathsf{v}}_j\}_{j=1}^m$ be unit normalized points sets of source and target images with same semantics. Then, we extract the inner and outer ellipsoids, $\underline{\mathcal{E}}$ and $\overline{\mathcal{E}}$, using (3.4) and (3.5), repetitively for $\mathcal{S}$ and $\mathcal{T}$. For $\hat{\mathsf{H}} \in SO(3)$, these two ellipsoids must satisfy,

$$\exists\lambda: \quad \lambda(\overline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal) \succeq \begin{bmatrix} \hat{\mathsf{H}} & 0 \\ 0 & 1 \end{bmatrix}^\intercal (\underline{\mathsf{Q}} - \mathsf{e}\mathsf{e}^\intercal) \begin{bmatrix} \hat{\mathsf{H}} & 0 \\ 0 & 1 \end{bmatrix}. \tag{12}$$

Remark 5.3 *LMI constraint for $\hat{\mathsf{H}} \in SO(3)$, involving only rotation with no scale, can be expressed similarly as in (10) by eliminating the scalar/scale variable α.*

We solve the problem of rotation estimation of purely rotating cameras using our Result 4.5 for (12), with the additional LMI constraint $\mathsf{I}_{4\times4} + \mathcal{L}(\hat{\mathsf{H}}) \succeq 0$.

6. Results

We present experiments for the problems described in Sec. 5, both on synthetic and real data. Our approach was implemented in MATLAB2017a using the Yalmip[3] toolbox and Mosek[4] as SDP solver. All experiments were carried out on an Intel Core i7 CPU 2.60GHz with 12GB RAM.

6.1. Synthetic Data

We show the general properties of our method for the two cases of Similarity Transform and Purely Rotating Camera problems. We proceed by synthetically generating points enclosed within an ellipsoid, as it can be seen in Fig. 2, representing each semantic region. These ellipsoids are called ***source ellipsoids***.

By applying an experiment-specific random transformation to the source ellipsoids, we obtain N ***target ellipsoids***, representing the ground-truth ellipsoid pair correspondences. To assess robustness, we generate test correspondences by adding different levels of noise to the point

[3] https://yalmip.github.io/
[4] https://www.mosek.com/

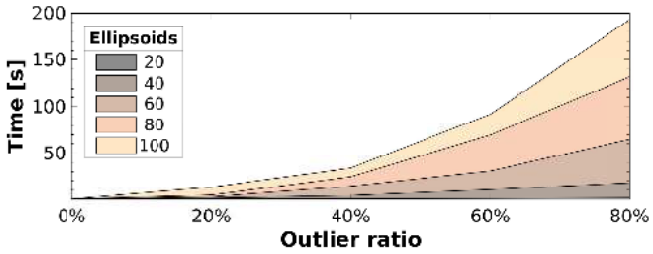

Figure 3: Runtime (synthetic data): with increasing number of ellipsoids and outlier ratio.

sets. In addition, we simulate outliers by adding a high amount of noise to a subset of these correspondences.

6.1.1 Similarity Transform

We begin by performing a series of experiments for the Similarity Transform problem. For all our experiments, we restrict $\alpha \in [0.2, 5.0]$, recall α in Theorem 5.1. In Fig. 3, we show runtimes for varying numbers of ellipsoidal correspondences and outlier ratios.

The metrics used for evaluating the quality of the results are: 3D root mean square error (RMSE), errors in rotation R, translation T, and scale S. For each experiment, we compute the errors $\Delta \mathsf{r} = ||\mathsf{r} - \mathsf{r}_{gt}||$, $\Delta \mathsf{t} = ||\mathsf{t} - \mathsf{t}_{gt}||$, and $\Delta s = ||s - s_{gt}||$. 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. The errors reported in Fig. 4 as ΔR, ΔT, and ΔS are the average values of 1000 experiments for the cases of 10, 20 and 30 ellipsoids.

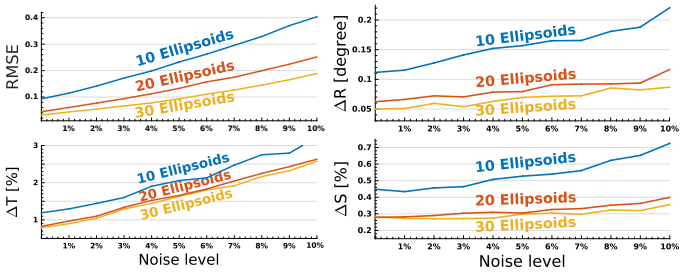

Figure 4: Errors Plots (synthetic data): 3D RMSE, Rotation (ΔR), Translation (ΔT), and Scale (ΔS) errors for different numbers of ellipsoids tested with increasing noise level.

In addition, Fig. 5 illustrates the behavior of the Branch-and-Bound during the exploration of the search space.

Local Refinement. Fig. 6 shows a comparison between a local method like ICP [6] that establish point correspondences, and our global method with ellipsoidal correspondences. The combination of both methods yields the lowest RMSE, because our method is only globally optimal with respect to ellipsoid correspondences. The ellipsoids-based representation allows for uncertainty during fitting and also

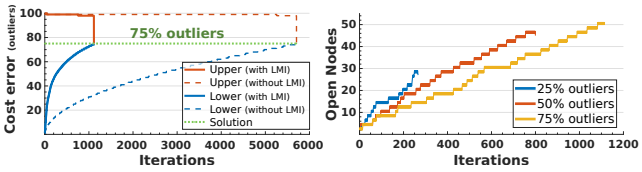

Figure 5: Branch-and-Bound (synthetic data). Plots with $N = 100$ ellipsoids. On the left, the bounds convergence are shown for an outlier ratio of 75 %, with and without the additional LMI constraint (eq. (10)). The runtime until convergence was 186 s and 890 s, respectively. Note the LMI constraints help to considerably reduce the number of iterations by pruning the search space. On the right, the plot shows how BnB scales with memory usage for outlier ratios of 25 %, 50 % and 75 %.

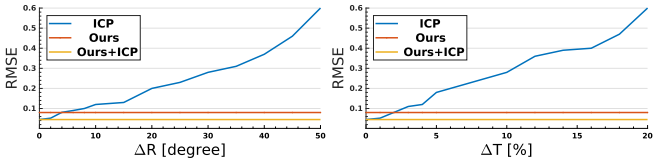

Figure 6: Local vs. Global (synthetic data): plot for $N = 20$ ellipsoids, each ellipsoid comprises 5 points for a total of 100 given points. The noise level was 2 %. We have generated different initializations by adding a perturbation in rotation (left) or translation (right) to the ground-truth alignment. Since our method is global, it is not affected by these different configurations. Note that local methods can be useful for further refinement of our results.

leads to less accurate registration results. In particular, the size ratio between the source and target ellipsoids essentially defines the noise level that our method will tolerate during the correspondence search (comparable to the inlier threshold for RANSAC). Therefore, a subsequent local refinement may further improve our estimation. Hence, our solution can be used to initialize local methods like ICP.

Moreover, we are targeting challenging registration tasks (Section 6.2) which cannot be handled by any variant of ICP (or other local methods) either due to the difference in data modalities, e.g. CAD model vs. scan, or due to the lack of a good initialization. Therefore, a global method that can make use of semantic labels is highly demanded for such tasks.

6.1.2 Purely Rotating Cameras

In Fig. 7, a comparison to other ***global methods*** is provided: Chin *et al*. [14], Bazin *et al*. [5], Speciale *et al*. [38]. The available open-source code of these methods was used without modifications. We conducted the experiments for the Purely Rotating Camera problem, which is the overlapping problem to all these methods, and whose runtime (for the exact same setup) is reported. The runtime comparison to [14] on homography estimation for purely rotating calibrated cameras is depicted in Fig. 7 as **A*-Search**.

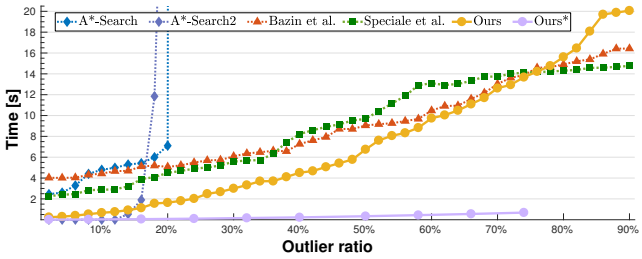

Figure 7: Global Methods (synthetic data): runtime comparison for the *purely rotating cameras* problem, among Chin *et al*. [14] (**A*-Search** and **A*-Search2**), Bazin *et al*. [5], Speciale *et al*. [38], Ours (50 Ellipsoids), Ours*(12 Ellipsoids, approx. $50/4$). Notice the latter case goes until less than $75\,\%$ since, passing this percentage, the number of points would not be sufficient to model the purely rotating camera.

For this method, it is also possible to use a linearized homography estimation and is therefore faster. The results of the linearized homography estimation code from [14] are also shown as **A*-Search2**. All methods were tested with $N = 50$ points on synthetic data, and 50 ellipsoids in our case. In addition, we also report the runtime for the case of only 12 ellipsoids (Ours*), since this is an approximate representation of 50 point correspondences, under the assumption that only 4 points are sufficient to model an ellipsoid.

6.2. Real Data

We tested our approach for a number challenging registration tasks: Day / Night, Outdoor / Indoor and Scan / CAD. Firstly, we explain the pipeline used to extract 3D ellipsoids for all three cases. Secondly, we describe the registration tasks in details; and finally, we present the qualitative and quantitative results.

Semantic Segmentation. We trained two deep convolutional neural networks [12] with modifications to the final class assignment. The networks operate in a two-phase approach, where overall categorical semantics are trained in one network (CatNet) and detailed object classes in a second network (DetNet). This step was necessary due to the imbalance of training data and to overcome final ambiguities. The final conditional random field optimization joins the two networks and provides a pixel-wise labeling [28].

The training data for the CatNet are taken from the Cityscape [19] and Mapillary [31] datasets, where coarse categories like *vegetation, building, ground, sky, etc.* are used. The DetNet combines likewise multiple datasets [11, 22, 36, 37, 41] for details on the buildings (windows, doors, balconies, etc.) and the previous datasets for less frequent classes like *car, pedestrian, traffic light, sidewalk, etc.*. All images are consolidated into a coherent label set, cropped to 321×321, and undergo standard augmentations for training

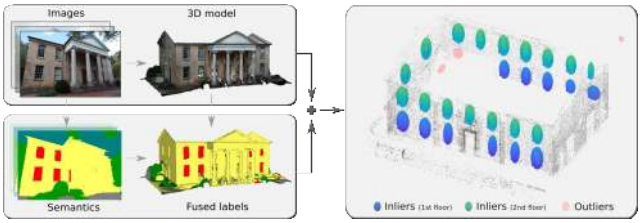

Figure 8: Ellipsoids pipeline: from images we obtain a 3D reconstruction model mesh, and pixel-wise semantically labeled images, which are later fused into the 3D mesh. Then, the 3D segmentation labels are clustered and filtered to obtain the ellipsoids.

50k iterations of the underlying ResNet-101 with a learning rate of $1e-4$ and a momentum of 0.9.

Likewise, the 2D semantic labeled images are projected onto the 3D surface mesh where the observations are fused over the triangles. Followed by a clustering method over the semantically labeled mesh vertices, we obtained a list of candidate ellipsoids which is further filtered to discard ellipsoids with too few labeled vertices inside (less than 200 points). In case all points belong to a plane, e.g. a window, the estimated ellipsoid will be too thin for the purpose of the feasibility test (5). In this case, we artificially inflate the ellipsoid by adding points in normal direction from the center. An example of this pipeline for extracting ellipsoids is shown in Fig. 8.

Day / Night Registration. We captured images of the same building during day and night, and then extracted the ellipsoids following the pipeline described in Fig. 8. As initial ellipsoid correspondences, one could create all-to-all correspondences within the same labels, but the problem becomes considerably big to handle. It is reasonable to assume some basic knowledge about the particular scene at hand. For instance, source ellipsoids lying on one plane (or wall) must also share a common plane on the target side. In the same fashion, we also incorporated the knowledge about different floor levels. Besides the windows, we also extracted ellipsoids for the doors and balconies, see Fig. 9. This greatly improves the search time by reducing the number of combinations. This is an example, where adding more semantic classes speeds up our method, as shown by the quantitative results in Table 1.

Outdoor / Indoor Registration. We used the data from [17] containing images of a theater from the outside and inside. These models are not connected and the goal is to align them by only using the information about the windows. The ellipsoids were obtained in the same manner as discussed before. Although the registration result shows a small difference in scale, our method is able to find a visually appealing solution which is comparable to the results reported in [17].

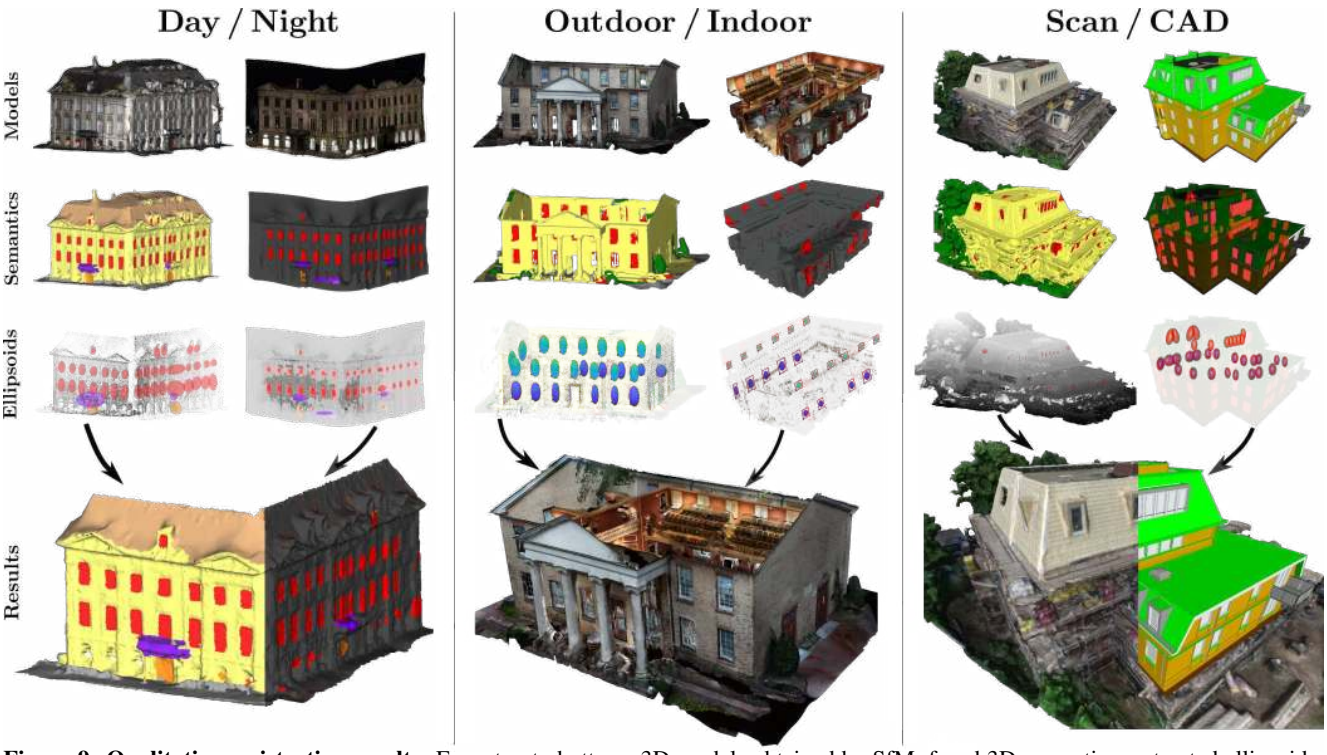

Figure 9: Qualitative registration results. From top to bottom: 3D models obtained by SfM; fused 3D semantics; extracted ellipsoids; and the aligned models. Note that we have two separated meshes for the Day / Night case. After the alignment we projected the semantic labels from the *night* mesh onto the *day* mesh. The Outdoor / Indoor case shows the merged colored SfMs instead of the semantics. For the Scan / CAD case, we depict a side-by-side comparison of the alignment models.

Scene	ΔR	ΔT	ΔS	$\lvert\zeta^*\rvert$ / N	Iterations	Time [sec]	Ours (RMSE)	Ours+ICP (RMSE)
Day / Night	$< 1^o$	$< 1\,\%$	$< 1\,\%$	30 / 202	108	117 s	0.19	0.08
Outdoor / Indoor	$< 1^o$	$< 2\,\%$	$< 3\,\%$	19 / 147	912	259 s	0.32	–
Scan / CAD	$< 3^o$	$< 2\,\%$	$< 2\,\%$	12 / 90	884	201 s	0.47	0.12

ΔR [degree]: rotation error. ΔT: translation error. ΔS: scale error. ζ^*: maximum consensus set. N: number of ellipsoid correspondences.

Table 1: Quantitative results. For each registration task, we show the rotation, translation, and scale error, followed by the number of inliers, total number of assignments, number of iterations, runtime and RMSE. Whenever possible, we applied ICP for refinement to obtain improved RMSE values (last column). For the Outdoor / Indoor case, ICP refinement was not possible due to the missing overlap between the models.

Scan / CAD Registration. We used the data from [1], which comprises a CAD model of a house, and weekly images captures by drones during the construction phase. The images used for the 3D reconstruction were from week number 30, when the building was still unfinished, see Fig. 9. This made the dataset highly challenging, besides the fact the data belongs to totally different modalities, namely: a 3D reconstructed model and a CAD model. Local methods like ICP fail on this dataset. We had difficulties to extract ellipsoid candidates for all floors and therefore concentrated on the second and fourth floor, where the windows were mostly visible from the drone. In order to account for noise and outliers we reduced the size of the inner ellipsoids. Then, we proceeded similarly as in the previous two tasks. The quantitative results for all three correspondence problems are reported in Table 1.

7. Conclusion

We proposed a novel method for finding correspondences between datasets with significantly different data statistics for which most traditional methods fail. Our approach incorporates semantic information of entire regions and accounts for noise and outliers in the low-level data by approximating them with ellipsoidal shapes. The correspondence problem is solved within a consensus maximization framework for which we derived necessary constraints that describe the assignments by ellipsoid inclusions. This leads to a generic framework to solve semantic registration problems in a globally optimal manner. Our experiments on both synthetic and real datasets demonstrate that our method is robust to large amounts of outliers and that it can be applied to a variety of challenging registration problems for which only specialized solutions exist in the literature.

References

[1] Schependomlaan dataset. https://github.com/openbimstandards/datasetschependomlaan, 2016.

[2] S. Alletto, G. Serra, and R. Cucchiara. Video registration in egocentric vision under day and night illumination changes. *Computer Vision and Image Understanding*, 157:274–283, 2017.

[3] M. Bansal and K. Daniilidis. Joint spectral correspondence for disparate image matching. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, pages 2802–2809, 2013.

[4] 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.

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

[6] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992.

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

[8] S. P. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory*, volume 15. SIAM, 1994.

[9] H. Bristow, J. Valmadre, and S. Lucey. Dense semantic correspondence where every pixel is a classifier. In *ICCV*, pages 4024–4031, 2015.

[10] D. Campbell, L. Petersson, L. Kneip, and H. Li. Globally-optimal inlier set maximisation for simultaneous camera pose and feature correspondence. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[11] J. Cech and R. Sara. Language of the structural models for constrained image segmentation grammars. Technical Report TN-eTRIMS-CMP-03-2007, CMP, Prague, 2007.

[12] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. In *ICLR*, 2015.

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

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

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

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

[17] A. Cohen, J. L. Schönberger, P. Speciale, T. Sattler, J.-M. Frahm, and M. Pollefeys. Indoor-outdoor 3d reconstruction alignment. In *European Conference on Computer Vision (ECCV)*, 2016.

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

[19] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *CVPR*, 2016.

[20] 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.

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

[22] R. Gadde, R. Marlet, and N. Paragios. Learning grammars for architecture-specific facade parsing. Technical report, INRIA, 2014.

[23] D. Glasner, S. N. P. Vitaladevuni, and R. Basri. Contour-based joint clustering of multiple segmentations. In *CVPR*, 2011.

[24] K. Han, R. S. Rezende, B. Ham, K. K. Wong, M. Cho, C. Schmid, and J. Ponce. Scnet: Learning semantic correspondence. *CoRR*, abs/1705.04043, 2017.

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

[26] D. C. Hauagge and N. Snavely. Image matching using local symmetry features. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pages 206–213, 2012.

[27] E. V. Haynsworth. On the schur complement. Technical report, DTIC Document, 1968.

[28] P. Kraehenbuehl and V. Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Proc. Neural Information Processing Systems (NIPS)*, 2011.

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

[30] D. Murray and A. Basu. Motion tracking with an active camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(5):449–459, May 1994.

[31] G. Neuhold, T. Ollmann, S. R. Bulo, and P. Kontschieder. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In *ICCV*, 2017.

[32] 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.

[33] D. P. Paudel, A. Habed, and L. V. Gool. Optimal transformation estimation with semantic cues. In *ICCV*, pages 4658–4667, 2017.

[34] F. Radenovic, J. L. Schönberger, D. Ji, J. Frahm, O. Chum, and J. Matas. From dusk till dawn: Modeling in the dark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 5488–5496, 2016.

[35] 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, August 2013.

[36] H. Riemenschneider, A. Bodis-Szomoru, J. Weissenberg, and L. Van Gool. Learning Where To Classify In Multi-View Semantic Segmentation. In *ECCV*, 2014.

[37] H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner, and H. Bischof. Irregular lattices for complex shape grammar facade parsing. In *CVPR*, 2012.

[38] P. Speciale, D. P. Paudel, M. R. Oswald, T. Kroeger, L. V. Gool, , and M. Pollefeys. Consensus maximization with linear matrix inequality constraints. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[39] R. Szeliski. Image alignment and stitching: A tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, January 2006.

[40] A. Toshev, J. Shi, and K. Daniilidis. Image matching via saliency region correspondences. In *CVPR*, 2007.

[41] R. Tyleček and R. Šára. Spatial pattern templates for recognition of objects with regular structure. In *Pattern Recognition (Proc. DAGM)*, 2013.

[42] F. Uhlig. A recurring theorem about pairs of quadratic forms and extensions: A survey. *Linear algebra and its applications*, 25:219–237, 1979.

[43] 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.

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

[45] 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.

[46] H. Zhou, T. Sattler, and D. W. Jacobs. Evaluating local features for day-night matching. In *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*, pages 724–736, 2016.

[47] J. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to-image translation. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 465–476, 2017.