

Robust and Optimal Sum-of-Squares-Based Point-to-Plane Registration of Image Sets and Structured Scenes

Danda Pani Paudel¹, Adlane Habed², Cédric Demonceaux¹, and Pascal Vasseur³

¹Le2i laboratory, University of Bourgogne Franche-Comté, CNRS, France

²ICube laboratory, University of Strasbourg, CNRS, France

³LITIS EA laboratory, University of Rouen, France,

{danda-pani.paudel, cedric.demonceaux}@u-bourgogne.fr, adlane.habed@icube.unistra.fr,
pascal.vasseur@univ-rouen.fr

Abstract

This paper deals with the problem of registering a known structured 3D scene and its metric Structure-from-Motion (SfM) counterpart. The proposed work relies on a prior plane segmentation of the 3D scene and aligns the data obtained from both modalities by solving the point-to-plane assignment problem. An inliers-maximization approach within a Branch-and-Bound (BnB) search scheme is adopted. For the first time in this paper, a Sum-of-Squares optimization theory framework is employed for identifying point-to-plane mismatches (i.e. outliers) with certainty. This allows us to iteratively build potential inliers sets and converge to the solution satisfied by the largest number of point-to-plane assignments. Furthermore, our approach is boosted by new plane visibility conditions which are also introduced in this paper. Using this framework, we solve the registration problem in two cases: (i) a set of putative point-to-plane correspondences (with possibly overwhelmingly many outliers) is given as input and (ii) no initial correspondences are given. In both cases, our approach yields outstanding results in terms of robustness and optimality.

1. Introduction

The emergence of affordable 3D sensors and high quality 2D cameras has triggered a growing interest in combining both imaging modalities. 3D sensors allow us to obtain faithful 3D scene models in the form of dense 3D point clouds while images can be used to extract texture information. High quality 3D models with mapped texture can be obtained provided the 2D and 3D sensors are registered in a common reference frame. The two imaging modalities are generally registered off-line and the 2D and 3D

sensors kept rigidly attached at all time during acquisition. Doing so may, however, be either impractical or impossible. On the one hand, suitable acquisition conditions for one sensor may not be adequate for the other (e.g. lighting conditions for cameras, surface orientation for 3D sensor, etc.) and, on the other hand, some application-specific requirements (e.g. camera on a drone and a 3D scanner on a vehicle) may altogether prohibit the sensors to be rigidly attached. When the cameras and the 3D sensor are free, reliable methods for registering the two modalities are highly desirable. This consists in establishing feature correspondences between the two modalities and estimating the rigid transformation aligning their respective reference frames.

Structure-from-Motion (SfM) techniques allow us to compute 3D point coordinates from pixel correspondences across images. It is thus tempting to regard the registration of 3D and 2D sensors as that of two 3D point sets: one set induced by the images and the other obtained from scanner measurements. Registering 3D point clouds is a well-studied problem. Most methods use the Iterative Closest Point (ICP) algorithm (or its variants) [31, 11, 24]. While ICP is a local method, recent work by Yang *et al.* [35] (Go-ICP) provides the very first globally optimal solution to same-scale point set registration. However, because SfM reconstructions suffer from a scale ambiguity, methods devised for registering same-scale data cannot be employed.

Most methods handling the scale ambiguity rely on establishing correspondences either between the 3D measurements obtained by both modalities or directly between scanned data and images [17, 5, 8]. The sought transformation parameters are then obtained either by minimizing the registration loss function or maximizing the consensus set of inliers. Note that Random Sample Consensus (RANSAC) [10] is the most widely used method for finding

the maximum set of inliers. Methods based on loss function minimization are more prone to outliers than their inlier-set-maximization counterparts [1]. Some methods exploit scene knowledge or the Manhattan World assumption. In this regard, methods have been devised based on line segment matching [17], target segmentation [30], repeated patterns detection [25], mutual information maximization [18], and extended Chamfer matching [37]. Registration methods that are based on establishing correspondences may be undermined by unreliable visual feature descriptors. Alternative methods, not establishing initial correspondences, have also been proposed [21, 7, 6]. The methods in [21, 7] use variants of the ICP algorithm and hence remain susceptible to partial scene overlap, scene occlusion, and high levels of outliers. The method in [6] employs a RANSAC-based inlier set maximization in which the scale problem is handled by an extension of the 4-point congruent sets algorithm.

As far as the problem of maximizing the set of inliers is concerned, RANSAC is non-deterministic and provides no guarantee with respect to the optimality of its solution. Globally optimal inlier set maximization methods [16, 1] have recently been proposed for problems that can be described using linear equations. However, extensions to problems with nonlinear equations [36] is problem-specific, difficult and may result in much more complicated (possibly numerically intractable) mathematical formulations. Note that a variety of methods for solving systems of nonlinear polynomial equations exist. While some are based on Gröbner bases or homotopy continuation [32], others use Sum-of-Squares (SoS) polynomial optimization [26, 15, 20]. However, such methods are dedicated to solving outlier-free nonlinear systems and dealing with outliers is carried out through RANSAC.

In this paper, we address the problem of registering the 3D scan and a set of images of a structured scene captured by calibrated cameras. Our assumption is that the scene is structured in the sense that it can be segmented into and represented by planes (or planar patches). Such representation is compact [2] and can also be useful for scene knowledge-based refinement methods [29]. The plane-based assumption is particularly valid when dealing with man-made environments, including (but not limited to) Manhattan World, urban and indoor scenes that are abundant with planes. In our approach, we seek the metric transformation relating the scene’s planes and SfM-induced 3D points. Note that point-to-plane registration methods are known to perform better than their point-to-point counterparts [27]. We rely on the fact that, under metric ambiguity, a point-to-plane assignment can be expressed as a second degree polynomial in scaled-quaternion and translation parameters. Our approach aims at maximizing the set of point-to-plane inliers with guaranteed optimality of the consensus set. The consensus set maximization methods [16, 1] discussed above

are not applicable because of the nonlinearity of the problem at hand. In our approach, we use the Branch-and-Bound (BnB) algorithmic paradigm to explore the scaled-quaternion and translation parameter space. As in [16, 1], we rely on establishing optimistic and pessimistic sets of point-to-plane inliers for pruning branches whose most optimistic sets are worse than the best pessimistic one. In this context, our contribution is threefold: (i) we propose a novel modeling of the point-to-plane (and point-to-patch) correspondence problem. Our modeling is based on a rigorous Sum-of-Squares polynomial optimization theory that allows us to identify, with certainty, point-to-plane mismatches within parameters’ bounds. This is used to derive optimistic sets of inliers where each point-to-plane assignment is considered independently from the others; (ii) we introduce SfM-specific constraints in our modeling, namely, a plane visibility criterion and optional vague constraints on the positions of the camera; (iii) based on our modeling and constraints, we propose a globally optimal approach for point-to-plane inlier set maximization in the presence of putative correspondences along with its non-combinatorial counterpart in the absence of point-to-plane correspondences.

2. The SoS theory

Definition 2.1 (SoS and PSD). *Let $\mathbb{R}[x]$ be the ring of polynomials in n variables, $x = (x_1, x_2, \dots, x_n)$, with real-valued coefficients. A polynomial $f(x) \in \mathbb{R}[x]$ is*

- *Positive Semi-Definite (PSD) (or nonnegative) if $f(x) \geq 0$ for all $x \in \mathbb{R}^n$;*

- *Sum-of-Squares (SoS) if there exist polynomials $f_i(x) \in \mathbb{R}[x]$ such that $f(x) = \sum_i f_i(x)^2$.*

A SoS is obviously always PSD and the converse is generally untrue. However, Hilbert [13] proved that, for some classes of polynomials including quadratic ones, a polynomial is PSD if and only if it is SoS. Checking whether a polynomial is PSD is NP-hard (though decidable) while checking whether a polynomial is SoS is computationally tractable using Semi-definite Programming (SDP) and employing the so-called Gram matrix of the polynomial.

Definition 2.2 (Gram matrix [22]). *Consider a polynomial $f(x) \in \mathbb{R}[x]$ of degree $2d$. Let $\mathbf{Z}_d(x)$ be the vector of monomials of $f(x)$ up to monomials of degree d . The matrix \mathbf{G} such that $f(x) = \mathbf{Z}_d(x)^\top \mathbf{G} \mathbf{Z}_d(x)$ is a Gram matrix of $f(x)$.*

Theorem 2.3 ([4, 22]). *A polynomial $f(x) \in \mathbb{R}[x]$ of degree $2d$ is SoS if and only if there exists a real symmetric positive semi-definite Gram matrix of $f(x)$.*

Note that since odd-degree polynomials cannot be SoS, only even-degree polynomials are concerned by such test. Checking for the existence of a positive semi-definite Gram matrix \mathbf{G} boils down to solving a Linear Matrix Inequality (LMI) feasibility problem. LMI feasibility can be efficiently solved using the interior-point algorithm [3]. Theorem 2.3

allows us to check whether a polynomial $f(\mathbf{x})$ is nonnegative for every $\mathbf{x} \in \mathbb{R}^n$. One is often interested in checking whether $f(\mathbf{x})$ is nonnegative in a semi-algebraic set \mathcal{K} defined by polynomials $g_i(\mathbf{x}) \in \mathbb{R}[x]$ such that

$$\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \geq 0, i = 1 \dots m\}. \quad (1)$$

This can be answered via the so-called *Positivstellensatz* (Psatz) [20]. The Psatz states that $f(\mathbf{x})$ is nonnegative on \mathcal{K} if there exist SoS polynomials $\sigma_v(\mathbf{x})$ such that

$$f(\mathbf{x}) = \sum_{v \in \{0,1\}^m} \sigma_v(\mathbf{x}) g_1(\mathbf{x})^{v_1} g_2(\mathbf{x})^{v_2} \dots g_m(\mathbf{x})^{v_m}. \quad (2)$$

Exploiting Psatz is difficult and may turn numerically intractable in practice because (2) requires 2^m SoS σ_v polynomials. Putinar [23] provides a much simpler Psatz under Archimedean conditions on the so-called quadratic module.

Definition 2.4 (Quadratic module [34]). *The quadratic module $M(g) = M(g_1, \dots, g_m) \subset \mathbb{R}[x]$ of polynomials $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x})$ is the set*

$$M(g) = \{\sigma_0(\mathbf{x}) + \sum_{i=1}^m \sigma_i(\mathbf{x}) g_i(\mathbf{x}) : \text{each } \sigma_i \text{ is SoS}\}. \quad (3)$$

Definition 2.5 (Archimedean [34]). *The quadratic module $M(g)$ of polynomials $g_1(\mathbf{x}), g_2(\mathbf{x}), \dots, g_m(\mathbf{x})$ is Archimedean if $N - \|\mathbf{x}\|^2 \in M(g)$ for some $N \in \mathbb{N}$.*

Theorem 2.6 (Putinar's Positivstellensatz [23]). *Assume the quadratic module $M(g)$ is Archimedean. If $f(\mathbf{x}) > 0$ on \mathcal{K} (defined by (1)), then $f(\mathbf{x}) \in M(g)$.*

3. SoS point-to-plane assignment conditions

We consider a set of two or more calibrated cameras observing a scene consisting of a set \mathcal{P} of at least four distinct planes in general positions. The scene has been scanned by a 3D sensor and segmented into these planes. We also consider the set \mathcal{Y} of seven or more points (lying on at least four distinct scene planes) whose projections are matched across two or more cameras. Let $\mathbf{y} \in \mathbb{R}^3$ be the SfM-induced [12] cartesian coordinate vector of a point $Y \in \mathcal{Y}$. The coordinates of the SfM-reconstructed points and those of the scene planes are represented in two distinct reference frames. A plane $\Pi \in \mathcal{P}$ is given by its normal 3-vector $\boldsymbol{\pi}$ and signed distance to the origin d . Because the image-induced reconstruction is metric, the transformation aligning the SfM-reconstructed points and the scanned scene is represented by a 3×3 scaled-rotation matrix \mathbf{Q} and a translation 3-vector \mathbf{t} . A quaternion representation with no enforcement of unit quaternion $\mathbf{q} = (z \ u \ v \ w)^\top$ is used to represent the scaled-rotation matrix \mathbf{Q} as follows:

$$\mathbf{Q} = \begin{bmatrix} z^2 + u^2 - v^2 - w^2 & 2uw - 2wz & 2uw + 2vz \\ 2uw + 2wz & z^2 - u^2 + v^2 - w^2 & 2vw - 2uz \\ 2uw - 2vz & 2vw + 2uz & z^2 - u^2 - v^2 + w^2 \end{bmatrix}.$$

Let $\mathcal{A} \subset \mathcal{Y} \times \mathcal{P}$ be the set of putative point-to-plane assignments (\times refers to the cartesian product) and $a = (Y, \Pi) \in \mathcal{A}$ is one such assignment. Furthermore, we denote by $\mathbf{x} \in \mathbb{R}^7$ the vector $\mathbf{x} = (\mathbf{q}^\top, \mathbf{t}^\top)^\top$ and let $f_a(\mathbf{x})$ be the polynomial in $\mathbb{R}[x]$ induced by a such that:

$$f_a(\mathbf{x}) := \boldsymbol{\pi}^\top (\mathbf{Q}\mathbf{y} + \mathbf{t}) - d. \quad (4)$$

If \mathbf{x} is the true registration parameter vector, then for every correct assignment $a \in \mathcal{A}$, $f_a(\mathbf{x}) = 0$. Our goal is to simultaneously estimate the registration parameters \mathbf{x} and associated set of correct point-to-plane assignments. Our approach is based on the BnB algorithmic paradigm and branching is carried out on the space of registration parameters \mathbf{x} . At each iteration, we are given parameter intervals, in the form of two vectors $\underline{\mathbf{x}}$ and $\bar{\mathbf{x}}$ in \mathbb{R}^7 whose respective entries \underline{x}_k and \bar{x}_k satisfy $\underline{x}_k \leq \bar{x}_k$ for $k = 1 \dots 7$. Although the full approach is detailed further in the paper, the idea is that such intervals are to be probed for point-to-plane potential assignments by attempting to solve the following problem:

Problem 3.1. *For a given $a \in \mathcal{A}$, is there a vector $\mathbf{x} \in \mathbb{R}^7$ satisfying $\underline{x}_k \leq x_k \leq \bar{x}_k$, $k = 1 \dots 7$ such that $f_a(\mathbf{x}) = 0$?*

In other words, one would like to know whether the polynomial crosses zero within the considered bounds. The point-to-plane assignment would then qualify as a potential inlier, i.e. possible correct assignment, within the considered bounds. This is however difficult to answer, unless the following alternative problem is considered:

Problem 3.2. *For a given $a \in \mathcal{A}$, is there a $\lambda_a \in \mathbb{R}$ such that $\lambda_a f_a(\mathbf{x}) > 0$ for every \mathbf{x} satisfying $\underline{x}_k \leq x_k \leq \bar{x}_k$ for $k = 1 \dots 7$?*

If $\lambda_a f_a(\mathbf{x}) > 0$, then the assignment a is definitely an outlier, i.e. incorrect assignment, within the bounds. Otherwise, it is a potential inlier. Indeed, if the question of Problem 3.2 is answered in the affirmative, the one of Problem 3.1 is answered in the negative: i.e. there exist no \mathbf{x} in the interval with which $f_a(\mathbf{x})$ crosses zero. Furthermore, one can rely on Putinar's Theorem 2.6 to solve Problem 3.2. To do so, assume we are given a set of polynomials $g_i(\mathbf{x})$ whose quadratic module $M(g)$ is Archimedean: if, for λ_a a scalar, $\lambda_a f_a(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathcal{K} = \{\mathbf{x} \in \mathbb{R}^7 : g_i(\mathbf{x}) \geq 0, i = 1 \dots m\}$, then $\lambda_a f_a(\mathbf{x}) \in M(g)$. Hence, there must exist SoS polynomials σ_i such that:

$$\lambda_a f_a(\mathbf{x}) - \sum_{i=1}^m \sigma_i(\mathbf{x}) g_i(\mathbf{x}) \text{ is SoS}. \quad (5)$$

Note that, in general, if (5) is satisfied, then $\lambda_a f_a(\mathbf{x})$ may not be necessarily positive in \mathcal{K} since \mathcal{K} could possibly be empty. However, so long as \mathcal{K} is not empty and σ_i SoS polynomials can be found, one is guaranteed that $\lambda_a f_a(\mathbf{x}) > 0$ everywhere in \mathcal{K} since $\sum_{i=1}^m \sigma_i(\mathbf{x}) g_i(\mathbf{x}) > 0$ in \mathcal{K} .

There are two main pending issues before one is able to use (5). First, one needs to find a set of polynomials $g_i(\mathbf{x})$,

representative of the parameter intervals, whose quadratic module $M(g)$ is Archimedean. Second, it is so far unclear how the σ_i SoS polynomials can be found. Let us explore now the first of these issues. Note that the Archimedean property is a matter of representation and the quadratic module of the set constructed from the linear interval constraints $x_k - \underline{x}_k \geq 0$ and $\bar{x}_k - x_k \geq 0$ is not Archimedean. In the following, we show that quadratic polynomial inequalities derived from such bound constraints yield an Archimedean quadratic module.

Proposition 3.3. *Consider the polynomials $g_k(x) = (x_k - \underline{x}_k)(\bar{x}_k - x_k)$, $k = 1 \dots 7$. The quadratic module $M(g)$ of these polynomials is Archimedean.*

Proof. As per Definition 2.5, for $M(g)$ to qualify as Archimedean, one must show that $N - \|x\|^2 \in M(g)$ for some $N \in \mathbb{N}$. In other words, there exist SoS $\sigma_0(x)$ and $\sigma_k(x)$, $k = 1 \dots 7$, such that $N - \sum_{k=1}^7 x_k^2 = \sigma_0(x) + \sum_{k=1}^7 \sigma_k(x)g_k(x)$. Equivalently, one needs to show that $N - \sum_{k=1}^7 x_k^2 - \sum_{k=1}^7 \sigma_k(x)g_k(x)$ is SoS. Upon expanding and factorizing the latter polynomial, we obtain $\sum_{k=1}^7 (\sigma_k(x) - 1)x_k^2 - \sum_{k=1}^7 \sigma_k(x)(\underline{x}_k + \bar{x}_k)x_k + (N + \sum_{k=1}^7 \sigma_k(x)\underline{x}_k\bar{x}_k)$. Using zero-degree SoS polynomials σ_k , i.e. nonnegative real scalars, one can always find $\sigma_k > 1$ and sufficiently large value of N such that this polynomial is always positive. Notice that the polynomial is quadratic in which case PSD and SoS are equivalent [13]. \square

Let us now consider the problem of checking whether or not (5) is SoS when considering the polynomials $g_k(x)$, $k = 1 \dots 7$ of Proposition 3.3. If so the assignment a is definitely an outlier within the bounds. If one knows beforehand that $\lambda_a f_a(x)$ must be positive, a sequence of $\sigma_k(x)$ of increasing degree can be used until a positivity certificate is obtained. However, for the problem at hand, when a set of $\sigma_k(x)$ of some degree fails to deliver such certificate, it is either because $\lambda_a f_a(x)$ indeed crosses zero (inlier) or the required degree for a positivity certificate has not been reached. The good news here is that, within a BnB search, the considered bound intervals $[\underline{x}, \bar{x}]$ get smaller and we show in the following that using nonnegative scalars σ_k rather than SoS polynomials of higher degree suffices. To see this, consider the following proposition:

Proposition 3.4. *Let $\hat{x} \in \mathbb{R}^7$ with known entries. The following statements are equivalent*

- (i) $\lambda_a f_a(\hat{x}) > 0$.
- (ii) \exists nonnegative scalars $\sigma_k \in \mathbb{R}$, $k = 1 \dots 7$:

$$\lambda_a f_a(x) + \sum_{k=1}^7 (x_k - \hat{x}_k)^2 \sigma_k > 0. \quad (6)$$

Proof. (ii) \implies (i) is straightforward. For (i) \implies (ii), consider $f_a(x)$'s Gram matrix G_f and G_x that of $\sum_{k=1}^7 (x_k - \hat{x}_k)^2$.

$\hat{x}_k)^2$. These matrices are defined by: $f_a(x) = x^T G_f x$ and $\sum_{k=1}^7 (x_k - \hat{x}_k)^2 = x^T G_x x$. Note that G_x is PSD and can be written as $G_x = U^T U$ with $U\hat{x} = 0$. The Gram matrix of the polynomial in (6) is then written as $\lambda_a G_f + U^T \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_7) U$. A direct application of Finsler's lemma [9] is that the latter matrix is positive-definite if and only if $\lambda_a \hat{x}^T G_f \hat{x} > 0$. This not only shows (i) \implies (ii) but also proves the equivalence. \square

We now state the following preliminary result:

Result 3.5. *Consider two vectors \underline{x} and \bar{x} in \mathbb{R}^7 whose respective entries \underline{x}_k and \bar{x}_k satisfy $\underline{x}_k \leq \bar{x}_k$ for $k = 1 \dots 7$. Let \mathcal{K}_b be the set*

$$\mathcal{K}_b = \{x \in \mathbb{R}^7 : g_k(x) := (x_k - \underline{x}_k)(\bar{x}_k - x_k) \geq 0\}. \quad (7)$$

If \exists a scalar λ_a and nonnegative scalars σ_k such that

$$\lambda_a f_a(x) - \sum_{k=1}^7 g_k(x) \sigma_k \quad (8)$$

is SoS, then $\lambda_a f_a(x) > 0$ for every $\underline{x}_k \leq x_k \leq \bar{x}_k$. In this case, the assignment a is guaranteed to be an outlier (a point-to-plane mismatch) within the considered bounds. Otherwise, a is a potential inlier. Furthermore, a consequence of Proposition 3.4 is that when $\bar{x}_k - \underline{x}_k$ tends towards zero, we are guaranteed that any outlier within the bound is detected. Indeed, this can be seen by noticing that when $\bar{x}_k = \underline{x}_k = \hat{x}_k$, polynomial (8) turns into (6).

Whether (8) is SoS can be tested by converting it into its corresponding Gram matrix LMI feasibility problem for the λ_a and σ_k indeterminates. Although the guarantee of identifying outliers using scalar σ_k multipliers is demonstrated with a zero-gap bound, in practice, outliers are detected very early in the process. As demonstrated in our experiments, the ability to detect outliers is improved with every size reduction of the investigated bounds. It may be tempting to use higher degree $\sigma_k(x)$ SoS polynomials to boost the process. However, this is unnecessary and yields slower performances compared to branching.

4. Registration

Our goal is to register a SfM-induced reconstruction and a plane-segmented scene. Unlike when dealing with 3D-3D registration, additional constraints emanating from the cameras can be exploited. Some may be implicit such as plane visibility, others, as vague camera locations, may be obtained from extra knowledge. In addition, when dealing with segmented scenes, one is given planar patches rather than infinite planes. Such additional constraints can augment the set \mathcal{K}_b derived from the bound constraints for earlier outlier detection. Note that adding new polynomial inequalities in \mathcal{K}_b has no effect on the Archimedean property of its quadratic module and Proposition 3.4 still holds.

Patches: Consider a scene plane Π and three or more planes Φ_k , not necessarily from the scene, orthogonal to it. The Φ_k planes must be chosen such that their intersection with Π defines a convex region on Π . The set of points on Π within this convex region is a patch. In practice, four such planes are adequate to represent meaningful patches in man-made environments. Each Φ_k is described by its normal vector ϕ_k and signed distance d_k . Let us denote by Φ the set $\{\Phi_k\}_{k=1}^4$ and let $\delta_k = \pm 1$ be the known sign, with respect to Φ_k , of a scanned point lying within the considered region. One can then identify outliers by checking whether $f_a(x)$ is positive everywhere within x 's bounds and in the set

$$\mathcal{K}_a^\Phi = \{x \in \mathbb{R}^7 : p_k(x) := (\phi_k^\top(Qy + t) - d_k)\delta_k \geq 0, \quad k = 1 \dots 4\}. \quad (9)$$

Plane visibility: Consider a point Y on a scene plane Π . If this point is imaged by two cameras, then these can only observe the same side of the plane: the one on which the point lies. In order for the cameras to observe the same side of the plane, their camera centers must lie on one side with respect to Π . Camera centers can easily be obtained from the SfM-calculated camera matrices: they are their right null space. Let C_k be the camera centers of $n \geq 2$ cameras with cartesian coordinates c_k . We define the set \mathcal{K}_Π^δ such that

$$\mathcal{K}_\Pi^\delta = \{x \in \mathbb{R}^7 : v_k(x) := (\pi^\top(Qc_k + t) - d)\delta \geq 0, \quad k = 1 \dots n\} \quad (10)$$

where $\delta = \pm 1$. We denote \mathcal{K}_Π^+ the set \mathcal{K}_Π^δ obtained using $\delta = +1$ and \mathcal{K}_Π^- otherwise. A given assignment a is a definite outlier if $f_a(x) > 0$ in \mathcal{K}_Π^+ and in \mathcal{K}_Π^- (in addition to patch and bounds conditions). Furthermore, planes for which $v_1(x)$ and $v_2(x)$ (for two cameras 1 and 2) always have opposite signs within x 's bounds cannot be assigned any points visible in those cameras. This would indicate that the plane always cuts the base-line of the two camera and cannot contain points visible in both cameras. Testing this can be carried out by checking, for $\delta = \pm 1$, whether

$$\begin{cases} \exists \sigma_k : v_1(x) - \sum_{k=1}^7 g_k(x)\sigma_k \text{ is SoS,} \\ \exists \sigma_k : -v_2(x) - \sum_{k=1}^7 g_k(x)\sigma_k \text{ is SoS.} \end{cases} \quad (11)$$

If for both values of δ , each polynomial in (11) is SoS, plane Π shall not be considered for assigning SfM points emanating from those cameras.

Camera bounds: A camera center C may lie within a box delimited by six planes in the set $\Psi = \{\Psi_k\}_{k=1}^6$ defined by their normal vectors ψ_k and signed distances d_k . Such information can be obtained from application-specific knowledge (GPS, moving vehicle, etc.). This knowledge can be used for further enforcing the search for point-to-plane outliers and turns very useful when no putative point-to-plane correspondences are initially known. Consider the cartesian coordinate vector c of the camera center and let

$$\mathcal{K}_c = \{x \in \mathbb{R}^7 : h_k(x) := (\psi_k^\top(Qc + t) - d_k)\delta_k \geq 0, \quad k = 1 \dots 6\} \quad (12)$$

where δ_k is the known sign, with respect to Φ_k , of any point within the considered box. If $h_k(x)$ are positive, the camera center is within the box. One can now test if $\lambda_a f_a(x) > 0$ whenever the camera center is in the box defined by \mathcal{K}_c .

Quaternions and scale: In the absence of scale, quaternion parameters demand that $q^\top q = 1$. When dealing with a scaled scene, the rotation is represented by a scaled quaternion matrix and one can only enforce that $q^\top q > 0$. It is understood that, in order to keep the problem numerically tractable via the Archimedean property, all registration parameters need to be bounded. The scale of the scene is no exception. When a better lower bound $\underline{s} > 0$ on the scale s is available, it is preferable to enforce that $q^\top q \geq \underline{s}$. This condition does not appear in the set \mathcal{K}_b and hence must be accounted for. Assuming the entries x_k , $k = 1 \dots 4$ of x correspond the quaternion parameters, we consider the set

$$\mathcal{K}_q = \{x \in \mathbb{R}^7 : q(x) := -\underline{s} + \sum_{k=1}^4 x_k^2 \geq 0\} \quad (13)$$

Furthermore, since both q and $-q$ yield the same rotation matrix, the initial lower bound of one of the quaternion parameters may arbitrarily be chosen nonnegative. The rest of the quaternion parameters may be initially bounded between $-\sqrt{\bar{s}}$ and $\sqrt{\bar{s}}$ where \bar{s} is the scale's upper bound.

We now state our main result:

Result 4.1. Assume we are given a putative point-to-plane assignment $a = (Y, \Pi) \in \mathcal{A}$, a patch on Π delimited by the planes in the set $\Phi = \{\Phi_k\}_{k=1}^4$, lower \underline{x} and upper \bar{x} bounds on the registration parameter vector x , bounds \underline{s} and \bar{s} on the scale of the scene, and (optionally) bounds defined by planes $\Psi = \{\Psi_k\}_{k=1}^6$ on the location of the camera centers of one (possibly more) camera. One would like to know whether or not the SfM-reconstructed point Y may lie on Π , while Π is visible by the cameras observing Y , within the patch Φ with registration parameters in the bounds \underline{x} and \bar{x} . In order to establish whether such assignment is possible, we consider the set

$$\mathcal{K} = \{x \in \mathbb{R}^7 : x \in \mathcal{K}_b \cap \mathcal{K}_a^\Phi \cap \mathcal{K}_\Pi^\delta \cap \mathcal{K}_c \cap \mathcal{K}_q\} \quad (14)$$

resulting from the intersection of all the sets defined by (7),(9),(10), (12) and (13). If there exist a scalar λ_a and nonnegative scalars σ_k , σ'_k , σ''_k , σ'''_k and σ such that

$$\lambda_a f_a(x) - \sum_{k=1}^7 g_k(x)\sigma_k - \sum_{k=1}^4 p_k(x)\sigma'_k - \sum_{k=1}^n v_k(x)\sigma''_k - \sum_{k=1}^6 h_k(x)\sigma'''_k - q(x)\sigma \quad (15)$$

is SOS, then $\lambda_a f_a(x) > 0$ in \mathcal{K} and the assignment a is a definite outlier. It is a potential inlier otherwise. Recall that this can be solved as a LMI feasibility problem.

Our registration approach is based on Result 4.1. We use in the following the term point-to-plane to refer to both point-to-plane and point-to-patch assignments. The goal of

the BnB algorithm is to estimate the registration parameters yielding the largest number of inliers. Our algorithm is provided a set of putative point-to-plane correspondences. In the absence of such correspondences, we consider every point to be putatively assigned to all the planes. A dynamically-built search tree, whose nodes are registration parameters' bounds, allows to explore the space of parameters. Given point-to-plane assignments and bounds on the registration parameters, the algorithm (see Algorithm 1) estimates the optimistic number of potential inliers using Result 4.1. A local method, a variant of the scaled-ICP algorithm [7], is used to obtain a pessimistic number of inliers for each given node. The local algorithm is started in the mid-values of the registration parameters' bounds. Its variation from [7] resides in constraining the registration solution to be within the investigated bounds in order to be representative of the node. We keep track of the highest (bestPessimistic in Algorithm 1) of the pessimistic number of potential inliers over all bound intervals. Any node whose optimistic number of inliers is worse than bestPessimistic is rejected. Otherwise, the node is qualified and branched along its longest edge resulting in two new nodes to be processed. The node corresponding to the bestPessimistic number of inliers is processed first. The algorithm terminates when no node has an optimistic number of inliers that is better than bestPessimistic.

Algorithm 1 Node processing

Input: bestPessimistic, registration param. bounds

Output: bestPessimistic

1. Count Optimistic no. of inliers using Result 4.1
 2. If Optimistic < bestPessimistic, reject the bounds.
 3. Count Pessimistic no. of inliers using local method.
 4. If (bestPessimistic < Pessimistic),
then bestPessimistic \leftarrow Pessimistic.
-

To qualify a point as an inlier, we distinguish two cases:
 1) Putative point-to-plane correspondences are provided: a point qualifies as a potential inlier if (15) is not proven SoS when assigned to the considered plane.
 2) No putative correspondences are provided: the point is considered a potential inlier as soon as (15) is not proven SoS when the point is assigned to one plane.

Discussion: In general, our method converges while the explored bounds are still quite large. The solution maximizing the inlier consensus set is the one returned by the local method. When the bounds are large enough, polynomials constructed from noisy data would still cross zero within the bounds allowing inliers, although affected by noise, to be accounted for. Therefore, the robustness to noise is more influenced by the local method than it is by the SoS tests. In our implementation, no special care was taken to further deal with noise when using SoS tests. However, in the case

of highly noisy data, the proposed SoS framework may allow to deal more efficiently with noise by incorporating an extra bounded variable ϵ (bounded by the allowed threshold), accounting for noise, in each point-to-plane assignment polynomial $f_a(x)$. In other words, though the assignment polynomial does not cross zero at the sought solution, $f_a(x) + \epsilon$ (for some value of ϵ) would. Furthermore, we have assumed throughout that the camera information fed to our algorithm is, to some extent, reliable. Should incorrect/noisy information about a camera be used, it may cause, especially with small camera bounding boxes, the registration to fail. In such cases, it is advised to include the camera-to-box constraints in the set of assignments to be accounted for when maximizing the consensus set.

5. Experiments

We conducted experiments with seven different benchmark real datasets shown in Figure 1 and whose details can be found in [14] and [28]. Our algorithm was implemented in MATLAB2014b and the SoS problems solved using the LMI Control Toolbox. All experiments were carried out on a 8GB RAM Pentium i7/3.40GHz. The SfM reconstructions and segmented scene planes were obtained using the openMVG Toolbox [19] and Hough Transform based plane detector [2]. For all the experiments, the initial bound on reconstruction scale was set to 0.2–5.0 (five times in scale in both directions). Four different error measurement metrics were defined to evaluate the registration quality: the RMS 3D error on normalized point sets, errors in rotation R , translation t , and scale s . For N experiments, these are defined as follows: $\Delta_R = (\frac{1}{3N} \sum_{i=1}^N \|r_i^* - r\|^2)^{1/2}$, $\Delta_T = (\frac{1}{N(\|t\|^2)} \sum_{i=1}^N \|t_i^* - t\|^2)^{1/2}$, $\Delta_S = (\frac{1}{N(s^2)} \sum_{i=1}^N (s_i^* - s)^2)^{1/2}$, where r is a vector obtained by stacking three rotation angles in degrees. The estimated variables are represented with * and variables without it are their ground truth.

5.1. Inlier set maximization with correspondences

The method was first tested for known putative correspondences where the synthetic inliers/outliers were generated under real data setups. No bounds on cameras were used in these experiments. To test the robustness, we varied the number of outliers up to 90% for Scene73 and compared the results against the linear 12-point RANSAC. Figure 2 shows that our method consistently detects 21 inliers for every experiment while RANSAC fails to detect the least number of required inliers starting from 45% of outliers. Note that the numbers of inliers reported here are true-positive inliers. Furthermore, our method does not detect any false positive inliers. Figure 4(left) shows the errors in rotation, translation, and scale for the same scene with various levels of outliers. The convergence graph of our method



Figure 1: Sample images from datasets (left to right): Scene23, Scene24, Scene27, Scene29, Scene73, Fountain, Herz-Jesu.

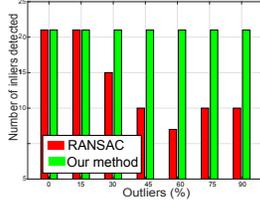
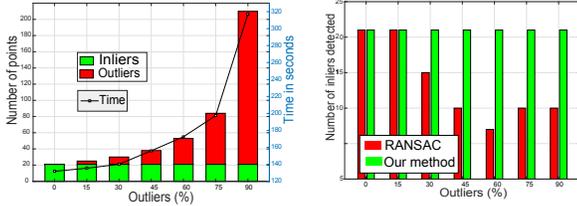


Figure 2: Experiments on Scene73 with correspondences and no camera bounds. Left: no. of processed points and time taken for various levels of outliers. Right: no. of detected inliers using RANSAC and our method.

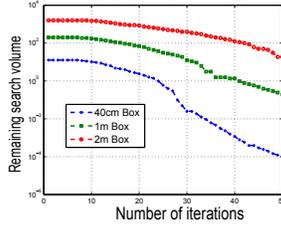
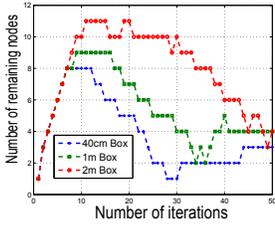


Figure 3: Experiments on Scene23 with correspondences, no camera bounds, and 50% outliers. Left: remaining nodes. Right: remaining volume.

	ΔR (degree)	ΔT (%)	ΔS (%)	3D error	Time (sec)
Scene23	0.785	1.75	0.21	0.0163	168.95
Scene73	1.263	4.63	1.68	0.0219	153.39
Fountain	0.524	1.21	0.53	0.0056	546.41

Table 1: Experiments with correspondences and no camera bounds: quantitative results obtained with 50% outliers.

with 50% outliers is shown in Figure 4(right) for Scene23, Scene73, and Fountain whose quantitative results are shown in Table 1. Figure 3 shows the evolution of the volume and the number of nodes remaining to be processed for the first 50 iterations on Scene23 with 50% outliers.

5.2. Inlier set maximization w/o correspondences

In the absence of initial correspondences, each point was assigned to all available planes. We conducted several experiments with bounded cameras by changing the number of bounded cameras and camera bounding box size. The number of iterations taken for these configurations are shown in Figure 6(left) for Scene23. The average time per iteration is 1.15sec. In the same figure, we also provide the number of iterations taken for the “with correspondences” case with 50% outliers and 2m camera bounding boxes. The case of a single bounded camera is equivalent to unbounded

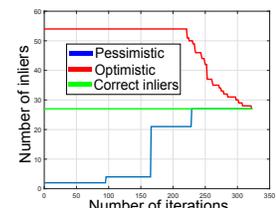
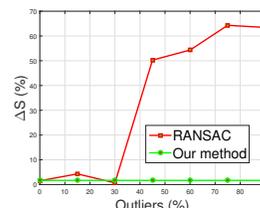
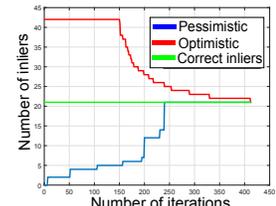
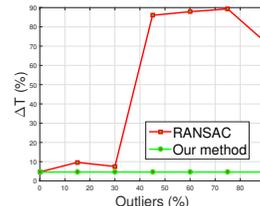
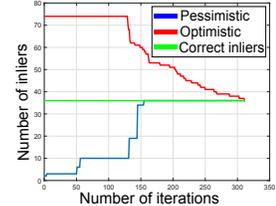
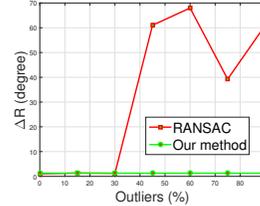


Figure 4: Experiments with correspondences and no camera bounds. Left: Error in rotation, translation, and scale for Scene73 vs. no. of outliers. Right: Convergence graph for 50% outliers (Top to bottom: Scene23, Scene73, Fountain).

cameras but bounded translation: plane visibility criterion cannot be used in this case. We recall that initial bounds on all the registration parameters are indispensable to ensure an Archimedean quadratic module of the constraints set and hence employ Putinar’s Psatz. Figure 6(right) shows the convergence graph, using Scene23, obtained with 3 1m-box bounded cameras. It also shows how the residual error on the registration parameters varies with the increase in the number of pessimistic inliers. The reported box size is for a normalized scene size of about 10 meters. In Figure 7, we report the results obtained on Scene23 (with 3 1m-box bounded cameras) using our method and a randomly started scaled ICP (RS-ICP) for 100 independent trials. In each trial, the scaled ICP was started at randomly picked registration parameter values satisfying bound and visibility constraints. The results show that, unlike RS-ICP which provides very large 3D errors, our method consistently detects the same number of inliers with the same 3D error.

The results of our method for all scenes (with their corresponding configurations) are summarized in Table 2. In

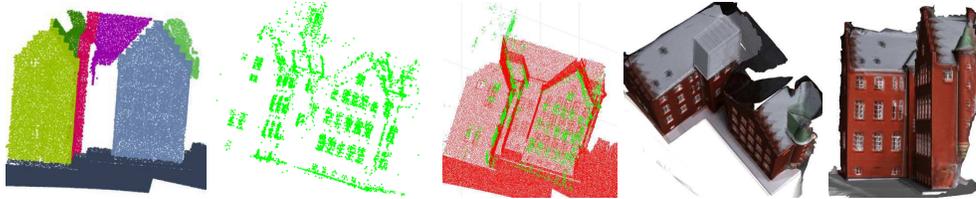


Figure 5: Left to right (Scene24): Segmented scene, reconstruction, registered point sets, two views of texture-mapped scene.

Scene	Points	Planes	Recon.	Rep.	Camera	Box	Inlier	ΔR	$\Delta T(\%)$	$\Delta S(\%)$	3D error	Iter	Time (sec)
Scene23	90	8	2.08134	0.8373	3	2m	41	3.4147	1.95	2.48	0.0619	482	599.738
Scene24	47	7	2.62791	0.8756	3	40cm	31	0.9591	2.31	2.09	0.0424	81	51.572
Scene27	49	4	1.61906	0.8127	8	50cm	20	2.5759	3.96	1.68	0.0131	133	141.837
Scene29	90	8	1.77408	0.9226	5	1m	45	2.9995	5.41	2.41	0.081	209	277.351
Scene73	71	8	3.21913	0.8654	5	1m	45	3.3463	4.78	2.73	0.0654	223	271.226
Fountain	29	7	0.81293	0.8495	4	40cm	11	2.8639	3.18	4.74	0.0570	102	55.730
Herz-Jesu	129	5	2.08134	0.6402	8	40cm	101	7.1958	4.02	1.99	0.0464	103	137.766

Table 2: Experiments w/o correspondences: quantitative results for seven different datasets.

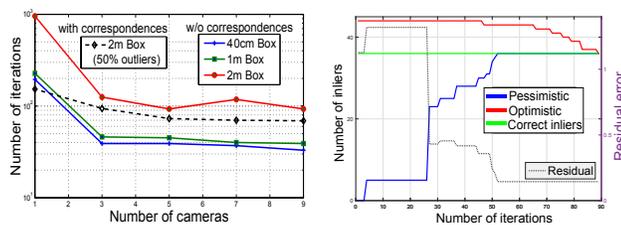


Figure 6: Experiments on Scene 23. Left: no. of iterations vs. no. of cameras. Right: Convergence graph for the case w/o correspondences with 3 1m-box bounded cameras.

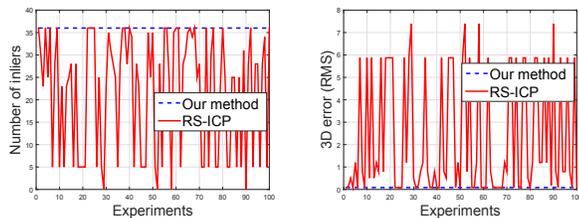


Figure 7: Experiments on Scene23 w/o correspondences with 3 1m-box bounded cameras (100 independent trials). Our method vs. randomly started scaled ICP (RS-ICP). Left: no. of inliers detected. Right: 3D RMS error.

Scene	Method	Time (sec)	ΔR	$\Delta T(\%)$	3D error
Fountain	RISAG	805.680	8.6825	14.08	0.3275
	Go-ICP	529.415	0.7225	1.63	0.0348
	Our method	55.730	2.8639	3.18	0.0570
Herz-Jesu	RISAG	160.064	17.6378	5.70	0.1830
	Go-ICP	31.254	3.2618	16.9	0.0725
	Our method	137.766	7.1958	4.02	0.0464

Table 3: Results using RISAG, Go-ICP and our method.

the reported parameters, Points, Planes, Iter, and Inlier represent their numbers. “Recon.” is the quality of the SfM reconstruction measured as the median reprojection error in pixels while “Rep.” is the fraction of the scene points represented by fitted planes. Observe that the registration quality depends upon the reconstruction quality, representation

factor, and the number and size of the camera boxes. For a qualitative evaluation, the results obtained for Scene24 are shown in Figure 5 along with the registered point sets and textured scene (after further refinement using [33]).

We also provide the results for two datasets obtained using RISAG [6], Go-ICP [35], and our method in Table 3. Our method was used without correspondences in the setting given in Table 2. Note that Go-ICP requires an Euclidean reconstruction, which was obtained by upgrading the metric reconstruction using ground truth measurements. Comparison of these methods may be unfair because each requires different initial conditions. Note that the poor performance of RISAG could be due to its RANSAC-driven nature (we used 10^4 RANSAC iterations).

6. Conclusion

We proposed a method for registering a 3D scan and a set of images of a scene represented by planes (or planar patches). Using Branch-and-Bound and SoS theory, we devised a robust and optimal method for inlier set maximization of point-to-plane correspondences. Although the problem is nonlinear and combinatorial, our method has provided outstanding results in terms of robustness. In the absence of initial assignments, the proposed method is non-combinatorial and can incorporate additional constraints that arise from plane visibility criterion and optional vague camera position constraints. The employed optimization framework has the potential to be efficiently applied to other nonlinear Computer Vision problems.

Acknowledgments

This research has been funded by the International Project NRF-ANR DrAACaR: ANR-11-ISO3-0003, the Regional Council of Bourgogne and European Regional Development Fund.

References

- [1] J. Bazin, H. Li, I. S. Kweon, C. Demonceaux, P. Vasseur, and K. Ikeuchi. A branch-and-bound approach to correspondence and grouping problems. In *PAMI*, pages 1565–1576, 2013. 2
- [2] D. Borrmann, J. Elseberg, K. Lingemann, and A. Nüchter. The 3d hough transform for plane detection in point clouds: A review and a new accumulator design. In *3D Research*, pages 32:1–32:13, 2011. 2, 6
- [3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004. 2
- [4] M. Choi, T. Lam, and B. Reznick. Sums of squares of real polynomials. *Proceedings of Symposia in Pure Mathematics*, 2(58):103–126, 1995. 2
- [5] S. Christy and R. Horaud. Iterative pose computation from line correspondences. In *CVIU*, pages 137–144, January 1999. 1
- [6] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno. Fully automatic registration of image sets on approximate geometry. *IJCV*, pages 91–111, March 2013. 2, 8
- [7] S. Du, N. Zheng, S. Ying, Q. You, and Y. Wu. An extension of the icp algorithm considering scale factor. In *ICIP*, volume 5, pages V–193. IEEE, 2007. 2, 6
- [8] L. Ferraz, X. Binefa, and F. Moreno-Noguer. Very fast solution to the pnp problem with algebraic outlier rejection. In *CVPR*, 2014. 1
- [9] P. Finsler. Über das vorkommen definitiver und semidefinitiver formen in scharen quadratischer formen. *Comment. Math. Helv.*, 9, pages 188–192, 1936/37. 4
- [10] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In *Commun. ACM*, pages 381–395, 1981. 1
- [11] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. In *BMVC*, pages 662–670, 2001. 1
- [12] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 3
- [13] D. Hilbert. Über die darstellung definitiver formen als summe von formen quadraten. In *Math. Ann.*, page 342350, 1888. 2, 4
- [14] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs. Large scale multi-view stereopsis evaluation. In *CVPR*, pages 406–413, 2014. 6
- [15] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. on Optimization*, pages 796–817, 2000. 2
- [16] H. Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *ICCV*, pages 1074–1080, 2009. 2
- [17] L. Liu and I. Stamos. Automatic 3d to 2d registration for the photorealistic rendering of urban scenes. In *CVPR*, pages 137–143, 2005. 1, 2
- [18] A. Mastin, J. Kepner, and J. W. Fisher III. Automatic registration of lidar and optical images of urban scenes. In *CVPR*, 2009. 2
- [19] P. Moulon, P. Monasse, and R. Marlet. Adaptive structure from motion with a contrario model estimation. In *ACCV*, pages 257–270, 2013. 6
- [20] P. A. Parrilo. Structured semidefinite programs and semi-algebraic geometry methods in robustness and optimization. Technical report, California Institute of Technology, 2000. 2, 3
- [21] D. P. Paudel, C. Demonceaux, A. Habed, and P. Vasseur. Localization of 2d cameras in a known environment using direct 2d-3d registration. In *ICPR*, pages 1–6, 2014. 2
- [22] V. Powers and T. Wörmann. An algorithm for sums of squares of real polynomials. *Journal of Pure and Applied Algebra*, 127(1):99 – 104, 1998. 2
- [23] M. Putinar. Positive polynomials on compact semi-algebraic sets. In *Indiana Univ. Math. J.*, pages 969–984, 1993. 3
- [24] S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *3DIM*, 2001. 1
- [25] G. Schindler, P. Krishnamurthy, R. Lubliner, Y. Liu, and F. Dellaert. Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In *CVPR*, pages 1–7, 2008. 2
- [26] G. Schweighofer and A. Pinz. Globally optimal o(n) solution to the pnp problem for general camera models. In *BMVC*, pages 1–10, 2008. 2
- [27] A. V. Segal, D. Haehnel, and S. Thrun. Generalized-icp. In *RSS*, 2009. 2
- [28] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *CVPR*, pages 1–8, 2008. 6
- [29] M. Tamaazousti, V. Gay-Bellile, S. N. Collette, S. Bourgeois, and M. Dhome. Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment. In *CVPR*, 2011. 2
- [30] A. Taneja, L. Ballan, and M. Pollefeys. City-scale change detection in cadastral 3d models using images. In *CVPR*, 2013. 2
- [31] A. B. Umberto Castellani. 3d shape registration. In *3D Imaging, Analysis, and Applications*, Springer, 2012. 1
- [32] J. Verschelde. Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Transactions on Mathematical Software (TOMS)*, 25(2):251–276, 1999. 2
- [33] P. Viola and W. M. Wells, III. Alignment by maximization of mutual information. *IJCV*, pages 137–154, September 1997. 8
- [34] S. Wagner. Archimedean quadratic modules: A decision problem for real multivariate polynomials. *Ph.D. thesis, Universität Konstanz*, August 2009. 3
- [35] J. Yang, H. Li, and Y. Jia. Go-icp: Solving 3d registration efficiently and globally optimally. In *ICCV*, pages 1457–1464, December 2013. 1, 8
- [36] J. Yang, H. Li, and Y. Jia. Optimal essential matrix estimation via inlier-set maximization. In *ECCV*, pages 111–126, 2014. 2
- [37] X. Zhang, G. Agam, and X. Chen. Alignment of 3d building models with satellite images using extended chamfer matching. In *CVPR Workshops*, pages 746–753, 2014. 2