Sparse Learning Approach to the Problem of Robust Estimation of Camera Locations

Arnak Dalalyan and Renaud Keriven
Université Paris Est, CERTIS, Ecole des Ponts ParisTech, Marne-la-Vallée, France
{dalalyan,keriven}@certis.enpc.fr

Abstract

In this paper, we propose a new approach—inspired by the recent advances in the theory of sparse learning— to the problem of estimating camera locations when the internal parameters and the orientations of the cameras are known. Our estimator is defined as a Bayesian maximum a posteriori with multivariate Laplace prior on the vector describing the outliers. This leads to an estimator in which the fidelity to the data is measured by the L_∞-norm while the regularization is done by the L_1-norm. Building on the papers [11, 15, 16, 14, 21, 22, 24, 18, 23] for L_∞-norm minimization in multiview geometry and, on the other hand, on the papers [8, 4, 7, 2, 1, 3] for sparse recovery in statistical framework, we propose a two-step procedure which, at the first step, identifies and removes the outliers and, at the second step, estimates the unknown parameters by minimizing the L_∞ cost function. Both steps are fairly fast: the outlier removal is done by solving one linear program (LP), while the final estimation is performed by a sequence of LPs. An important difference compared to many existing algorithms is that for our estimator it is not necessary to specify neither the number nor the proportion of the outliers.

1. Introduction

In the present paper, we are concerned with the structure and motion problem of multiview geometry. This problem, that have received a great deal of attention by the computer vision community in last decade, consists in recovering a set of 3D points (structure) and a set of camera matrices (motion), when only 2D images of the aforementioned 3D points by some cameras are available. Throughout this work we assume that the internal parameters of cameras as well as their orientations are known. Thus, only the locations of camera centers and 3D points are to be estimated. In solving the structure and motion problem by state-of-the-art methods, it is customary to start by establishing correspondences between pairs of 2D data points. We will assume in the present study that these point correspondences have been already established.

From a heuristical standpoint, it is useful to think of the structure and motion problem as an inverse problem. In fact, if $\mathcal{O}$ denotes the operator that takes as input the set of 3D points and the set of cameras, and produces as output the 2D images of the 3D points by the cameras, then the task of the structure and motion problem is to invert the operator $\mathcal{O}$. It is clear that the operator $\mathcal{O}$ is not injective: two different inputs may result in the same output. Fortunately, in many situations (for example, when for each pair of cameras there are at least five 3D points in general position that are seen by these cameras [20]), there is only a small number of inputs, up to an overall similarity transform, having the same image by $\mathcal{O}$. In such cases, the solutions to the structure and motion problem can be found using algebraic arguments. The solutions obtained by the algebraic approach are in general very sensitive to the noise in the data. Thus, very often, there is no input that could have generated the observed output because of the noise in the measurements. A natural approach to cope with such situations consists in searching for the input providing the closest possible output to the observed data.

In the problems where the output space of the operator $\mathcal{O}$ is not one-dimensional, as it is the case in the structure and motion problem, the choice of the metric used to measure the "closeness" is of high importance. A standard approach [12] consists in measuring the distance between two elements of the output space in the Euclidean L_2-norm. In the structure and motion problem with more than two cameras, this leads to a hard non-convex optimization problem. A particularly elegant way of circumventing the non-convexity issues inherent to the use of L_2-norm consists in replacing it by the L_∞-norm. This approach has been developed in a series of recent papers (cf. [11, 14, 21, 22, 24] and the references therein). It has been shown that, for some problems of structure and motion estimation, the use of the L_∞-norm results in a pseudoconvex minimization, which can be performed very efficiently using, for example, the iterative bisection method that solves a convex program at

each iteration.

The purpose of the present work is to introduce a new procedure of estimation in presence of noise and outliers. Our procedure is derived as a maximum a posteriori (MAP) estimator under uniformly distributed random noise and a sparsity favoring prior on the vector of outliers. Interestingly, this study bridges the work on the robust estimation in multiview geometry and the well developed theory of sparse recovery in statistical learning theory [4, 7, 5, 17]. Related work on outlier removal using L_∞-techniques can be found in [9, 24, 15, 18]. We refer the interested reader to the paper [24] for a comprehensive, precise mathematical background on the outlier identification problem in multiview geometry.

The rest of the paper is organized as follows. The next section gives the precise formulation of the translation estimation problem that constitutes an example of a problem to which the presented methodology can be applied. A brief review of the L_∞-norm minimization algorithm consisting in sequential optimization is presented in Section 3. In Section 4, we introduce the statistical framework and derive a new estimation procedure as a MAP estimator. The main result on the accuracy of this procedure is stated and proved in Section 5, while Section 6 contains some numerical experiments. The methodology of our study is summarized in Section 7.

2. Translation estimation and triangulation

Let us start by presenting a problem of multiview geometry to which our approach can be successfully applied: the problem of translation estimation and triangulation in the case of known rotations. For rotation estimation algorithms, we refer the interested reader to [19, 10] and the references therein. Let $\mathbf{P}_i^*$, $i = 1, \ldots, m$, be a sequence of m cameras that are known up to a translation. Recall that a camera is characterized by a 3×4 matrix $\mathbf{P}$ with real entries that can be written as $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$, where $\mathbf{K}$ is an invertible 3×3 camera calibration matrix, $\mathbf{R}$ is a 3×3 rotation matrix and $\mathbf{t} \in \mathbb{R}^3$. We will refer to $\mathbf{t}$ as the translation of the camera $\mathbf{P}$. We can thus write $\mathbf{P}_i^* = \mathbf{K}_i[\mathbf{R}_i|\mathbf{t}_i^*]$, $i = 1, \ldots, m$. For a set of unknown 3D points $\mathbf{U}_j^* \in \mathbb{P}^3$, $j = 1, \ldots, n$, we are given the images—contaminated by a noise— of each $\mathbf{U}_j^*$ by some cameras $\mathbf{P}_i^*$. Thus, we have at our disposal the measurements

$$\mathbf{x}_{ij} = \frac{1}{\mathbf{e}_3^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^*} \begin{bmatrix} \mathbf{e}_1^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^* \\ \mathbf{e}_2^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^* \end{bmatrix} + \boldsymbol{\xi}_{ij}, \quad \begin{matrix} j = 1, \ldots, n, \\ i \in I_j, \end{matrix} \quad (1)$$

where $\mathbf{e}_\ell$, $\ell = 1, 2, 3$, stands for the unit vector of $\mathbb{R}^3$ having one as the ℓth coordinate and I_j is the set of indices of cameras for which the point $\mathbf{U}_j^*$ is visible. We will assume that the set $\{\mathbf{U}_j^*\}$ does not contain points at infinity. Therefore, we can write $\mathbf{U}_j^* = [\mathbf{X}_j^{*\mathsf{T}}|1]^\mathsf{T}$ for some $\mathbf{X}_j^* \in \mathbb{R}^3$ and for every $j = 1, \ldots, n$.

We are now in a position to state the problem of translation estimation and triangulation in the context of multiview geometry. It consists in recovering the 3-vectors $\{\mathbf{t}_i^*\}$ (translation estimation) and the 3D points $\{\mathbf{X}_j^*\}$ (triangulation) from the noisy measurements $\{\mathbf{x}_{ij}; j = 1, \ldots, n; i \in I_j\} \subset \mathbb{R}^2$. We use the notation $\boldsymbol{\theta}^* = (\mathbf{t}_1^{*\mathsf{T}}, \ldots, \mathbf{t}_m^{*\mathsf{T}}, \mathbf{X}_1^{*\mathsf{T}}, \ldots, \mathbf{X}_n^{*\mathsf{T}})^\mathsf{T} \in \mathbb{R}^{3(m+n)}$, which is the vector of interest.

Remark 1 (Cheirality). *If the point $\mathbf{U}_j^*$ is in front of the camera $\mathbf{P}_i^*$, then $\mathbf{e}_3^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^* \geq 0$. Furthermore, we will assume that none of the true 3D points $\mathbf{U}_j^*$ lies on the principal plane of a camera $\mathbf{P}_i^*$. This assumption implies that $\mathbf{e}_3^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^* > 0$ so that the quotients $\mathbf{e}_\ell^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^* / \mathbf{e}_3^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^*$, $\ell = 1, 2$, are well defined.*

Remark 2 (Identifiability). *The parameter $\boldsymbol{\theta}$ is, in general, not identifiable from the measurements $\{\mathbf{x}_{ij}\}$. In fact, for every $\alpha \neq 0$ and for every $\mathbf{t} \in \mathbb{R}^3$, the parameters $\{\mathbf{t}_i^*, \mathbf{X}_j^*\}$ and $\{\alpha(\mathbf{t}_i^* - \mathbf{R}_i \mathbf{t}), \alpha(\mathbf{X}_j^* + \mathbf{t})\}$ generate the same measurements. To cope with this issue, we assume that $\mathbf{t}_1^* = \mathbf{0}_3$ and that $\min_{i,j} \mathbf{e}_3^\mathsf{T} \mathbf{P}_i^* \mathbf{U}_j^* = 1$. Thus, in what follows we assume that $\mathbf{t}_1^*$ is removed from $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^* \in \mathbb{R}^{3(m+n-1)}$. Further assumptions ensuring the identifiability are given below.*

3. Estimation by L_∞-minimization

This section presents results on the estimation of $\boldsymbol{\theta}$ based on the reprojection error minimization. This material is essential for understanding the results that are at the core of the present work. In what follows, for every $s \geq 1$, we denote by $\|\mathbf{x}\|_s$ the L_s-norm of a vector $\mathbf{x}$, *i.e.* $\|\mathbf{x}\|_s^s = \sum_j |x_j|^s$ if $\mathbf{x} = (x_1, \ldots, x_d)^\mathsf{T}$ for some $d > 0$. As usual, we extend this to $s = +\infty$ by setting $\|\mathbf{x}\|_\infty = \max_j |x_j|$.

3.1. Estimation by L_∞ cost minimization

A classical method [12] for estimating the parameter $\boldsymbol{\theta}$ is based on minimizing the sum of the reprojection errors squared. This amounts to defining $\widehat{\boldsymbol{\theta}}$ as a minimizer of the cost function $\mathcal{C}_{2,2}(\boldsymbol{\theta}) = \sum_{i,j} \|\mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})\|_2^2$, where $\mathbf{x}_{ij}(\boldsymbol{\theta}) := [\mathbf{e}_1^\mathsf{T} \mathbf{P}_i \mathbf{U}_j^*; \mathbf{e}_2^\mathsf{T} \mathbf{P}_i \mathbf{U}_j^*]^\mathsf{T} / \mathbf{e}_3^\mathsf{T} \mathbf{P}_i \mathbf{U}_j^*$ is the 2-vector that we would obtain if $\boldsymbol{\theta}$ were the true parameter. It can also be written as

$$\mathbf{x}_{ij}(\boldsymbol{\theta}) = \left[\frac{\mathbf{e}_1^\mathsf{T} \mathbf{K}_i (\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)}{\mathbf{e}_3^\mathsf{T} \mathbf{K}_i (\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)}; \frac{\mathbf{e}_2^\mathsf{T} \mathbf{K}_i (\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)}{\mathbf{e}_3^\mathsf{T} \mathbf{K}_i (\mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i)} \right]^\mathsf{T}. \quad (2)$$

The minimization of $\mathcal{C}_{2,2}$ is a hard nonconvex problem. In general, it does not admit closed-form solution and the existing iterative algorithms may often get stuck in local minima. An ingenious idea to overcome this difficulty has been proposed in recent years [11, 13]. It is based on the minimization of the L_∞ cost function

$$\mathcal{C}_{\infty,2}(\boldsymbol{\theta}) = \max_{j=1,\ldots,n} \max_{i \in I_j} \|\mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})\|_2. \quad (3)$$

Note that the substitution of the L_2-cost function by the L_∞-cost function has been proved to lead to improved algorithms in other estimation problems as well, cf., *e.g.*, [6]. This cost function has the advantage of having all its sub-level sets convex. This property ensures that all minima of $\mathcal{C}_{\infty,2}$ form a convex set and that an element of this set can be computed by solving a sequence of SOCPs [14], *e.g.* by the bisection algorithm presented below. This result readily generalizes to the family of cost functions

$$\mathcal{C}_{\infty,s}(\boldsymbol{\theta}) = \max_{j=1,\ldots,n} \max_{i \in I_j} \|\mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})\|_s, \quad s \geq 1. \quad (4)$$

For $s = 1$ and $s = +\infty$, the minimization of $\mathcal{C}_{\infty,s}$ can be recast in a sequence of LPs.

3.2. The bisection algorithm

We briefly describe an algorithm computing $\widehat{\boldsymbol{\theta}}_s \in \arg\min_{\boldsymbol{\theta}} \mathcal{C}_{\infty,s}(\boldsymbol{\theta})$ for any prespecified $s \geq 1$. The minimization is carried out over the set of all vectors $\boldsymbol{\theta}$ satisfying the cheirality condition. Let us introduce the residuals $\mathbf{r}_{ij}(\boldsymbol{\theta}) = \mathbf{x}_{ij} - \mathbf{x}_{ij}(\boldsymbol{\theta})$ that can be represented as

$$\mathbf{r}_{ij}(\boldsymbol{\theta}) = \left[\frac{\mathbf{a}_{ij1}^\mathsf{T}\boldsymbol{\theta}}{\mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta}}; \frac{\mathbf{a}_{ij2}^\mathsf{T}\boldsymbol{\theta}}{\mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta}} \right]^\mathsf{T}, \quad (5)$$

for some vectors $\mathbf{a}_{ij\ell}, \mathbf{c}_{ij} \in \mathbb{R}^2$. Furthermore, as presented in Remark 2, the cheirality conditions imply the set of linear constraints $\mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta} \geq 1$. Thus, the problem of computing $\widehat{\boldsymbol{\theta}}_s$ can be rewritten as

$$P_s: \quad \min_{\{\boldsymbol{\theta}, \gamma\}} \gamma \quad \text{s.t.} \quad \begin{cases} \|\mathbf{r}_{ij}(\boldsymbol{\theta})\|_s \leq \gamma, \\ \mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta} \geq 1. \end{cases} \quad (6)$$

Note that the inequality $\|\mathbf{r}_{ij}(\boldsymbol{\theta})\|_s \leq \gamma$ can be replaced by $\|\mathtt{A}_{ij}^\mathsf{T}\boldsymbol{\theta}\|_s \leq \gamma \mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta}$ with $\mathtt{A}_{ij} = [\mathbf{a}_{ij1}; \mathbf{a}_{ij2}]$. Although P_s is not a convex problem, its solution can be well approximated by solving a sequence of convex feasibility problems of the form

$$P_{s,\gamma}: \quad \text{find} \quad \boldsymbol{\theta} \quad \text{s.t.} \quad \begin{cases} \|\mathtt{A}_{ij}^\mathsf{T}\boldsymbol{\theta}\|_s \leq \gamma \mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta}, \\ \mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta} \geq 1, \end{cases} \quad (7)$$

with fixed $\gamma \geq 0$.

Remark 3. *Any solution of (7) is defined up to a scaling factor larger than one,* i.e., *if $\boldsymbol{\theta}$ solves (7) then it is also the case for $\alpha\boldsymbol{\theta}$ with any $\alpha \geq 1$. In order to fix this scaling factor, one can redefine $\boldsymbol{\theta}$ by setting it equal to $\boldsymbol{\theta}/(\min_{ij} \mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta})$.*

Given a small positive number ϵ controlling the accuracy of approximation, the bisection algorithm reads as follows:

Step 1: Compute a $\widehat{\boldsymbol{\theta}}$ satisfying the cheirality conditions, *e.g.*, by solving a linear feasibility problem.
Step 2: Set $\gamma_l = 0$ and $\gamma_u = \mathcal{C}_{\infty,s}(\widehat{\boldsymbol{\theta}})$.
Step 3: Set $\gamma = (\gamma_l + \gamma_u)/2$.
Step 4: If $P_{s,\gamma}$ has no solution, set $\gamma_l = \gamma$. Otherwise, replace the current value of $\widehat{\boldsymbol{\theta}}$ by a solution to $P_{s,\gamma}$ and set $\gamma_u = \mathcal{C}_{\infty,s}(\widehat{\boldsymbol{\theta}})$.
Step 5: If $\gamma_u - \gamma_l < \epsilon$, then assign to $\widehat{\boldsymbol{\theta}}_s$ the current value of $\widehat{\boldsymbol{\theta}}$ and terminate. Otherwise, go to Step 3.

4. Robust estimation by linear programming

This and the next sections contain the main theoretical contribution of the present work. We start with the precise formulation of the statistical model and define a maximum a posteriori (MAP) estimator.

4.1. The statistical model

Let us first observe that, in view of (1) and (5), the model we are considering can be rewritten as

$$\left[\frac{\mathbf{a}_{ij1}^\mathsf{T}\boldsymbol{\theta}^*}{\mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta}^*}; \frac{\mathbf{a}_{ij2}^\mathsf{T}\boldsymbol{\theta}^*}{\mathbf{c}_{ij}^\mathsf{T}\boldsymbol{\theta}^*} \right]^\mathsf{T} = \boldsymbol{\xi}_{ij}, \quad j = 1, \ldots, n; \ i \in I_j. \quad (8)$$

Let $N = 2\sum_{j=1}^n I_j$ be the total number of measurements and let $M = 3(n + m - 1)$ be the size of the vector $\boldsymbol{\theta}^*$. Let us denote by $\mathtt{A}$ (resp. $\mathtt{C}$) the $M \times N$ matrix formed by the concatenation of the column-vectors $\mathbf{a}_{ij\ell}$ (resp. $\mathbf{c}_{ij}^1$). Similarly, let us denote by $\boldsymbol{\xi}$ the N-vector formed by concatenating the vectors $\boldsymbol{\xi}_{ij}$. In these notation, Eq. (8) is equivalent to

$$\mathbf{a}_p^\mathsf{T}\boldsymbol{\theta}^* = (\mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}^*)\boldsymbol{\xi}_p, \quad p = 1, \ldots, N. \quad (9)$$

This equation defines the statistical model in the case where there is no outlier or, equivalently, all the measurements are inliers. In order to extend this model to cover the situation where some outliers are present in the measurements, we assume that there is another vector, $\boldsymbol{\omega}^* \in \mathbb{R}^N$, such that $\omega_p^* = 0$ if the pth measurement is an inlier and

$$\mathbf{a}_p^\mathsf{T}\boldsymbol{\theta}^* = \omega_p^* + (\mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}^*)\boldsymbol{\xi}_p, \quad p = 1, \ldots, N. \quad (10)$$

It is convenient to write these equations in a matrix form:

$$\mathtt{A}^\mathsf{T}\boldsymbol{\theta}^* = \boldsymbol{\omega}^* + \text{diag}(\mathtt{C}^\mathsf{T}\boldsymbol{\theta}^*)\boldsymbol{\xi}, \quad (11)$$

where, as usual, for every vector $\mathbf{v}$, $\text{diag}(\mathbf{v})$ is the diagonal matrix having the components of $\mathbf{v}$ as diagonal entries.

Statement of the problem: *Given the matrices $\mathtt{A}$ and $\mathtt{C}$, estimate the parameter-vector $\boldsymbol{\beta}^* = [\boldsymbol{\theta}^{*\mathsf{T}}; \boldsymbol{\omega}^{*\mathsf{T}}]^\mathsf{T}$ based on the following prior information:*

---

[1]To get a matrix of the same size as $\mathtt{A}$, in the matrix $\mathtt{C}$ each column is duplicated two times.

C_1: *Eq. (11) holds with some small noise vector $\boldsymbol{\xi}$,*

C_2: $\min_p \mathbf{c}_p^\mathsf{T} \boldsymbol{\theta}^* = 1$ *(cf. the discussion preceding Eq. (6)),*

C_3: *$\boldsymbol{\omega}^*$ is sparse, i.e., only a small number of coordinates of the vector $\boldsymbol{\omega}^*$ are different from zero.*

4.2. Sparsity prior and MAP estimator

To derive an estimator of the parameter $\boldsymbol{\beta}^*$, we place ourselves in the Bayesian framework. To this end, we impose a probabilistic structure on the noise vector $\boldsymbol{\xi}$ and introduce a prior distribution on the unknown vector $\boldsymbol{\beta}$. Since the noise $\boldsymbol{\xi}$ represents the difference (in pixels) between the measurements and the true image points, it is naturally bounded and, generally, does not exceeds the level of a few pixels. In view of this boundedness, it is reasonable to assume that the components of $\boldsymbol{\xi}$ are uniformly distributed in some compact set of $\mathbb{R}^2$, centered at the origin. We assume in what follows that the subvectors ξ_{ij} of $\boldsymbol{\xi}$ are uniformly distributed in the square $[-\sigma, \sigma]^2$ and are mutually independent. This implies that all the coordinates of $\boldsymbol{\xi}$ are independent.

Now, we turn to the choice of the prior. Since the only information on $\boldsymbol{\theta}^*$ is that the cheirality and the identifiability constraints should be satisfied, we define the prior on $\boldsymbol{\theta}$ as the uniform distribution on the polytope $\mathcal{P} = \{\boldsymbol{\theta} \in \mathbb{R}^M : \mathsf{C}^\mathsf{T}\boldsymbol{\theta} \geq 1\}$, where the inequality is understood componentwise. (For simplicity, we assume in this discussion that $\mathcal{P}$ is bounded.) The density of this distribution is $p_1(\boldsymbol{\theta}) \propto \mathbf{1}_\mathcal{P}(\boldsymbol{\theta})$, where $\propto$ stands for the proportionality relation and $\mathbf{1}_\mathcal{P}(\boldsymbol{\theta}) = 1$ if $\boldsymbol{\theta} \in \mathcal{P}$ and 0 otherwise.

The task of choosing a prior on $\boldsymbol{\omega}$ is more delicate in that it should reflect the information that $\boldsymbol{\omega}$ is sparse. The most natural prior would be the one having a density which is a decreasing function of the L_0-norm of $\boldsymbol{\omega}$, i.e., of the number of its nonzero coefficients. It is however well-known that the computation of estimators based on this type of priors is NP-hard. A very powerful approach for overcoming this difficulty—extensively studied in the statistical literature during the past decade ([4, 7, 1] and the references therein) —relies on using the L_1-norm instead of the L_0-norm. Following this idea, we define the prior on $\boldsymbol{\omega}$ by the density $p_2(\boldsymbol{\omega}) \propto f(\|\boldsymbol{\omega}\|_1)$, where f is some decreasing function[2]. Assuming in addition that $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ are independent, we get the following prior on $\boldsymbol{\beta}$:

$$\pi(\boldsymbol{\beta}) = \pi(\boldsymbol{\theta}; \boldsymbol{\omega}) \propto \mathbf{1}_\mathcal{P}(\boldsymbol{\theta}) \cdot f(\|\boldsymbol{\omega}\|_1). \qquad (12)$$

Theorem 1. *Assume that the noise $\boldsymbol{\xi}$ has independent entries which are uniformly distributed in $[-\sigma, \sigma]$ for some $\sigma > 0$, then the MAP estimator $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\theta}}^\mathsf{T}; \widehat{\boldsymbol{\omega}}^\mathsf{T}]^\mathsf{T}$ based on the prior π defined by Eq. (12) is the solution of the opti-*

---

[2]The most common choice is $f(x) = e^{-x}$ corresponding to the multivariate Laplace density.

mization problem:

$$LP_\sigma : \quad \min \|\boldsymbol{\omega}\|_1 \quad s.t. \begin{cases} |\mathbf{a}_p^\mathsf{T}\boldsymbol{\theta} - \omega_p| \leq \sigma\mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}, \ \forall p \\ \mathbf{c}_p^\mathsf{T}\boldsymbol{\theta} \geq 1, \ \forall p. \end{cases}$$
(13)

Proof. Under the probabilistic assumption made on the vector $\boldsymbol{\xi}$, the conditional probability density $f_{\boldsymbol{\beta}}$ of the data $\mathbb{X} = \{\mathbf{x}_{ij}, j = 1, \ldots, n; i \in I_j\}$ given the parameter-vector $\boldsymbol{\beta}$ is $f_{\boldsymbol{\beta}}(\mathbb{X}) \propto \prod_{p=1}^N \mathbf{1}_{\{|\mathbf{a}_p^\mathsf{T}\boldsymbol{\theta} - \omega_p| \leq \mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}\}}$. Consequently, the posterior probability density is given by

$$f_{\boldsymbol{\beta}|\mathbb{X}}(\boldsymbol{\beta}) \propto \prod_{p=1}^N \mathbf{1}_{\{|\mathbf{a}_p^\mathsf{T}\boldsymbol{\theta} - \omega_p| \leq \mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}\}} \mathbf{1}_\mathcal{P}(\boldsymbol{\theta}) \cdot f(\|\boldsymbol{\omega}\|_1). \quad (14)$$

Since f is a decreasing function, it is obvious that the set of most likely values w.r.t. this posterior distribution coincides with the set of solutions to the problem LP_σ. $\qquad\square$

Remark 4 (Condition C_2). *One easily checks that any solution of LP_σ satisfies condition C_2. Indeed, if for some solution $\widehat{\boldsymbol{\beta}}$ it were not the case, then $\min_p \mathbf{c}_p^\mathsf{T}\widehat{\boldsymbol{\theta}} > 1$. Therefore, $\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}/\min_p \mathbf{c}_p^\mathsf{T}\widehat{\boldsymbol{\theta}}$ would satisfy the constraints of LP_σ and $\tilde{\boldsymbol{\omega}}$ would have a smaller L_1-norm than $\widehat{\boldsymbol{\omega}}$, which is in contradiction with the fact that $\widehat{\boldsymbol{\beta}}$ solves LP_σ.*

Remark 5 (The role of σ). *In the definition of $\widehat{\boldsymbol{\beta}}$, σ is a free parameter that can be interpreted as the level of separation of inliers from outliers. In fact, the proposed algorithm implicitly assumes that all the measurements $\mathbf{x}_{ij}$ for which $\|\boldsymbol{\xi}_{ij}\|_\infty > \sigma$ are outliers, while all the others are treated as inliers.*

In the case when σ is unknown, a reasonable way of acting is to impose a prior distribution on the possible values of σ and to define the estimator $\widehat{\boldsymbol{\beta}}$ as a MAP estimator based on the prior incorporating the uncertainty on σ. When there are no outliers and the prior on σ is decreasing, this approach leads to the estimator minimizing the L_∞ cost function (see Section 3). In the presence of outliers, the role of the shape of the prior on σ, as well as the shape of the function f (see Eq. (15)), becomes more important for the definition of the estimator. This is an interesting point for future investigation.

4.3. Two-step procedure

Building on the arguments presented in the previous subsection, we introduce the following two-step algorithm.

Input: $\{\mathbf{a}_p, \mathbf{c}_p; p = 1, \ldots, N\}$ and σ.

Step 1: Compute $[\widehat{\boldsymbol{\theta}}^\mathsf{T}; \widehat{\boldsymbol{\omega}}^\mathsf{T}]^\mathsf{T}$ as a solution to LP_σ, cf. (13).

Step 2: Define $J = \{p : \widehat{\omega}_p = 0\}$ and apply the bisection algorithm to the reduced data set $\{\mathbf{x}_p; p \in J\}$.

Two observations are in order. First, when applying the bisection algorithm at Step 2, we can use $\mathcal{C}_{\infty,s}(\widehat{\boldsymbol{\theta}})$ as the initial value of γ_u. The second observation is that a better way of acting would be to minimize the weighted L_1-norm of $\boldsymbol{\omega}$, where the weight assigned to ω_p is inversely proportional to the depth $\mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}^*$. Since $\boldsymbol{\theta}^*$ is unknown, a reasonable strategy consists in adding a step in between Step 1 and Step 2, which performs the weighted minimization with weights $\{(\mathbf{c}_p^\mathsf{T}\widehat{\boldsymbol{\theta}})^{-1}; p = 1, \ldots, N\}$.

5. Accuracy of estimation

Let us introduce some additional notation. Recall the definition of $\mathcal{P}$ and set $\partial\mathcal{P} = \{\boldsymbol{\theta} : \min_p \mathbf{c}_p^T\boldsymbol{\theta} = 1\}$ and $\Delta\mathcal{P} = \{\boldsymbol{\theta} - \boldsymbol{\theta}' : \boldsymbol{\theta}, \boldsymbol{\theta}' \in \partial\mathcal{P}\}$. For every subset of indices $J \subset \{1, \ldots, N\}$, we denote by $\mathtt{A}_J$ the $M \times N$ matrix obtained from $\mathtt{A}$ by replacing the columns that have an index outside J by zero. Furthermore, let us define

$$\delta_S = \max_{|J| \leq S} \sup_{\mathbf{g} \in \Delta\mathcal{P}} \frac{\|\mathtt{A}_J^\mathsf{T}\mathbf{g}\|_2}{\|\mathtt{A}^\mathsf{T}\mathbf{g}\|_2}, \qquad \forall S \leq N, \quad (15)$$

where $|J|$ is the cardinal of J and the convention $0/0 = 0$ is used. One easily checks that δ_S is well defined and is less than or equal to 1 for every S. Moreover, the mapping $S \mapsto \delta_S$ is nondecreasing.

Assumption A: *The real number λ defined by $\lambda = \min_{\mathbf{g} \in \Delta\mathcal{P}} \|\mathtt{A}^\mathsf{T}\mathbf{g}\|_2 / \|\mathbf{g}\|_2$ is strictly positive.*

It is worth mentioning that Assumption A is necessary for identifying the parameter vector $\boldsymbol{\theta}^*$. In fact, if we consider the case without outliers, $\widehat{\boldsymbol{\omega}}^* = \mathbf{0}$, and if Assumption A is not fulfilled, then there is a vector $\mathbf{g} \in \Delta\mathcal{P}$ such that $\mathtt{A}^\mathsf{T}\mathbf{g}$ is much smaller than $\mathbf{g}$. That is, given the matrices $\mathtt{A}$ and $\mathtt{C}$, there are two vectors $\boldsymbol{\theta}^1$ and $\boldsymbol{\theta}^2$ satisfying $\min_p \mathbf{c}_p\boldsymbol{\theta}^j = 1$, $j = 1, 2$, and such that $\mathtt{A}^\mathsf{T}(\boldsymbol{\theta}^1 - \boldsymbol{\theta}^2)$ is small while $\boldsymbol{\theta}^1 - \boldsymbol{\theta}^2$ is not. Therefore, if eventually $\boldsymbol{\theta}^1$ is the true parameter vector satisfying $\mathtt{C}_1$ and $\mathtt{C}_3$, then $\boldsymbol{\theta}^2$ satisfies these conditions as well but is substantially different from $\boldsymbol{\theta}^1$. As a consequence, the true vector cannot be accurately estimated from matrices $\mathtt{A}$ and $\mathtt{C}$. Note also that a simple sufficient condition for Assumption A to be fulfilled is that the smallest eigenvalue of $\mathtt{A}\mathtt{A}^\mathsf{T}$ is strictly positive.

5.1. The noise free case

To evaluate the quality of estimation, we first place ourselves in the case where $\sigma = 0$. The estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}^*$ is then given as a solution to the optimization problem

$$\min \|\boldsymbol{\omega}\|_1 \quad \text{over } \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\omega} \end{bmatrix} \text{ s.t. } \begin{cases} \mathtt{A}^\mathsf{T}\boldsymbol{\theta} = \boldsymbol{\omega} \\ \mathtt{C}^\mathsf{T}\boldsymbol{\theta} \geq 1 \end{cases}. \quad (16)$$

Theorem 2. *Let Assumption A be fulfilled and let $\delta_S + \delta_{2S} < 1$ for some $S \in \{1, \ldots, N\}$. Then, for some constant C_0, it holds:*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C_0 \|\boldsymbol{\omega}^* - \boldsymbol{\omega}_S^*\|_1, \quad (17)$$

where $\boldsymbol{\omega}_S^*$ stands for the vector $\boldsymbol{\omega}^*$ with all but the S-largest entries set to zero. In particular, if $\boldsymbol{\omega}^*$ has no more than S nonzero entries, then the estimation is exact: $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^*$.

Proof. We set $\mathbf{h} = \boldsymbol{\omega}^* - \widehat{\boldsymbol{\omega}}$ and $\mathbf{g} = \boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}$. It follows from Remark 4 that $\mathbf{g} \in \Delta\mathcal{P}$. From now on, $\mathbf{h}_T$ stands for the vector equal to $\mathbf{h}$ on an index set T and zero elsewhere. Let T_0 (resp. T_1) denote the index set corresponding to the locations of S largest entries[3] of $\boldsymbol{\omega}^*$ (resp. $\mathbf{h}_{T_0^c}$, where T_0^c is the complementary set of T_0). To proceed with the proof, we need the following auxiliary result.

Lemma 1. *Let $\mathbf{v} \in \mathbb{R}^d$ be some vector and let $S \leq d$ be a positive integer. If we denote by T the indices of S largest entries of the vector $|\mathbf{v}|$ (the absolute value should be understood componentwise), then $\|\mathbf{v}_{T^c}\|_2 \leq S^{-1/2}\|\mathbf{v}\|_1$.*

Proof of Lemma 1. Let us denote by T_1 the index set of S largest entries of $|\mathbf{v}_{T^c}|$, by T_2 the index set of next S largest entries of $|\mathbf{v}_{T^c}|$, and so on. By triangle inequality, one has $\|\mathbf{v}_{T^c}\|_2 \leq \sum_{j \geq 1} \|\mathbf{v}_{T_j}\|_2$. On the other hand, one easily checks that $|v_\ell|^2 \leq |v_\ell| \cdot \|\mathbf{v}_{T_{j-1}}\|_1/S$ for every $\ell \in T_j$ with the convention $T_0 = T$. This implies that $\|\mathbf{v}_{T_j}\|_2^2 \leq \|\mathbf{v}_{T_j}\|_1\|\mathbf{v}_{T_{j-1}}\|_1/S$, for every $j \geq 1$. After taking the square root of these inequalities and summing up over j, we get the desired result in view of the obvious inequality $\|\mathbf{v}_{T_j}\|_1 \leq \|\mathbf{v}_{T_{j-1}}\|_1$. $\qquad\square$

Applying Lemma 1 to the vector $\mathbf{v} = \mathbf{h}_{T_0^c}$ and to the index set $T = T_1$, we get

$$\|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 \leq S^{-1/2}\|\mathbf{h}_{T_0^c}\|_1. \quad (18)$$

On the other hand, summing up the inequalities $\|\mathbf{h}_{T_0^c}\|_1 \leq \|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0^c}\|_1 + \|\boldsymbol{\omega}_{T_0^c}^*\|_1$ and $\|\boldsymbol{\omega}_{T_0}^*\|_1 \leq \|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0}\|_1 + \|\mathbf{h}_{T_0}\|_1$, and using the relation $\|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0}\|_1 + \|(\boldsymbol{\omega}^* - \mathbf{h})_{T_0^c}\|_1 = \|\boldsymbol{\omega}^* - \mathbf{h}\|_1 = \|\widehat{\boldsymbol{\omega}}\|_1$, we get

$$\|\mathbf{h}_{T_0^c}\|_1 + \|\boldsymbol{\omega}_{T_0}^*\|_1 \leq \|\widehat{\boldsymbol{\omega}}\|_1 + \|\boldsymbol{\omega}_{T_0^c}^*\|_1 + \|\mathbf{h}_{T_0}\|_1. \quad (19)$$

Since $\boldsymbol{\beta}^*$ satisfies the constraints of the optimization problem (16) a solution of which is $\widehat{\boldsymbol{\beta}}$, we have $\|\widehat{\boldsymbol{\omega}}\|_1 \leq \|\boldsymbol{\omega}^*\|_1$. This inequality, in conjunction with (18) and (19), implies

$$\|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 \leq S^{-1/2}\|\mathbf{h}_{T_0}\|_1 + 2S^{-1/2}\|\boldsymbol{\omega}_{T_0^c}^*\|_1$$
$$\leq \|\mathbf{h}_{T_0}\|_2 + 2S^{-1/2}\|\boldsymbol{\omega}_{T_0^c}^*\|_1, \quad (20)$$

---

[3] in absolute value

where the last line follows from the Cauchy-Schwartz inequality. Once again using the fact that both $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ satisfy the constraints of (16), we get $\mathbf{h} = \mathtt{A}^\mathsf{T}\mathbf{g}$. Therefore,

$$
\begin{aligned}
\|\mathbf{h}\|_2 &\leq \|\mathbf{h}_{T_0 \cup T_1}\|_2 + \|\mathbf{h}_{(T_0 \cup T_1)^c}\|_2 \\
&\leq \|\mathbf{h}_{T_0 \cup T_1}\|_2 + \|\mathbf{h}_{T_0}\|_2 + 2S^{-1/2}\|\boldsymbol{\omega}^*_{T_0^c}\|_1 \\
&= \|\mathtt{A}^\mathsf{T}_{T_0 \cup T_1}\mathbf{g}\|_2 + \|\mathtt{A}^\mathsf{T}_{T_0}\mathbf{g}\|_2 + 2S^{-1/2}\|\boldsymbol{\omega}^*_{T_0^c}\|_1 \\
&\leq (\delta_{2S} + \delta_S)\|\mathtt{A}^\mathsf{T}\mathbf{g}\|_2 + 2S^{-1/2}\|\boldsymbol{\omega}^*_{T_0^c}\|_1 \\
&= (\delta_{2S} + \delta_S)\|\mathbf{h}\|_2 + 2S^{-1/2}\|\boldsymbol{\omega}^*_{T_0^c}\|_1. \quad (21)
\end{aligned}
$$

Since $\boldsymbol{\omega}^*_{T_0^c} = \boldsymbol{\omega}^* - \boldsymbol{\omega}_S$, the last inequality yields

$$
\|\mathbf{h}\|_2 \leq \left(2S^{-1/2}/(1 - \delta_S - \delta_{2S})\right)\|\boldsymbol{\omega}^* - \boldsymbol{\omega}^*_S\|_1. \quad (22)
$$

To complete the proof, it suffices to observe that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 &\leq \|\mathbf{g}\|_2 + \|\mathbf{h}\|_2 \leq \lambda^{-1}\|\mathtt{A}\mathbf{g}\|_2 + \|\mathbf{h}\|_2 \\
&= (\lambda^{-1} + 1)\|\mathbf{h}\|_2 \leq C_0\|\boldsymbol{\omega}^* - \boldsymbol{\omega}^*_S\|_1, \quad (23)
\end{aligned}
$$

by virtue of inequality (22). □

We emphasize that the constant C_0 is rather small. For example, if $\delta_S + \delta_{2S} = 0.5$, then it can be deduced from the proof of Theorem 2 that $\max(\|\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}^*\|_2, \|\mathtt{A}^\mathsf{T}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\|_2) \leq (4/\sqrt{S})\|\boldsymbol{\omega}^* - \boldsymbol{\omega}^*_S\|_1$.

5.2. The noisy case

The assumption $\sigma = 0$ is an idealization of the reality that has the advantage of simplifying the mathematical derivations. While such a simplified setting is useful for conveying the main ideas behind the proposed methodology, it is of major practical importance to discuss the extensions to the more realistic *noisy* model. Because of space limitation, we will state the result which is the analogue of Theorem 2 in the case $\sigma \neq 0$.

Theorem 3. *Let us denote by* $\widehat{\boldsymbol{\xi}}$ *the vector of estimated residuals satisfying* $\mathtt{A}^\mathsf{T}\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\omega}} + \mathrm{diag}(\mathtt{C}^\mathsf{T}\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\xi}}$. *If for some* $\epsilon > 0$ *we have* $\max(\|\mathrm{diag}(\mathtt{C}^\mathsf{T}\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\xi}}\|_2; \|\mathrm{diag}(\mathtt{C}^\mathsf{T}\boldsymbol{\theta}^*)\boldsymbol{\xi}\|_2) \leq \epsilon$, *then*

$$
\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_2 \leq C_0\|\boldsymbol{\omega}^* - \boldsymbol{\omega}^*_S\|_1 + C_1\epsilon \quad (24)
$$

where C_0 *and* C_1 *are numerical constants.*

The proof of this result is very similar to that of Theorem 2 and therefore is left to the interested reader.

6. Numerical illustration

We implemented the algorithm in MatLab, using the SeDuMi package for solving linear programs [26]. Since in the present setting the matrices A and C are sparse—each column of these matrices contains at most 6 non-zero
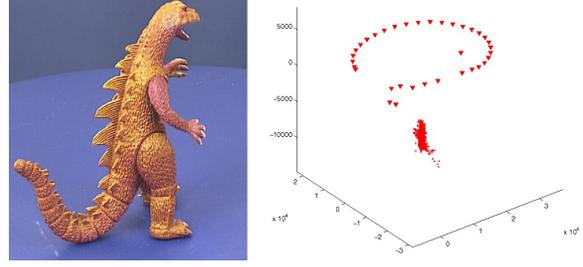

Figure 1. *Left: the first image of the database. Right: the positions of cameras and the scene points estimated by L_∞-norm minimization.*

entries—the execution times are reasonably small even for large data sets.

To demonstrate the proposed methodology, we applied our algorithm of robust estimation to the well known dinosaur sequence[4]. This sequence consists of 36 images of a dinosaur on a turntable, see Fig. 1 for the first image. The 2D image points which are tracked across the image sequence and the projection matrices of 36 cameras are provided as well. For solving the problem of translation estimation and triangulation, we make use only of the first three columns of the projection matrices.

There are at least two factors making the analysis of the dinosaur data difficult. The first one is the size of the data set: there are 16.432 image points corresponding to 4.983 real world points. This entails that there are more than 15.000 unknown parameters. The second factor is the presence of outliers which causes the failure of the original bisection algorithm. As shown in Fig. 1, the estimated camera centers are not on the same plane and it is difficult to recognize the dinosaur from the scatter plot of the estimated 3D points. Furthermore, the maximal reprojection error in this example is equal to 63 pixel.

We ran our procedure of robust estimation on this data set with $\sigma = 0.5$ pixel. We applied the following rule for detecting outliers: if $|\omega_p/\mathbf{c}_p^\mathsf{T}\boldsymbol{\theta}|$ is larger than $\sigma/4$, then the pth measurement is considered as an outlier and is removed. The corresponding 3D scene point is also removed if, after the step of outlier removal, it was seen by only one camera. This resulted in removing 1.306 image points and 297 scene points. The plots of Fig. 2 show the estimated camera centers and estimated scene points. We see, in particular, that the camera centers are almost coplanar. Note that in this example, the second step of the procedure described in Section 4.3 does not improve on the estimator computed at the first step. Thus, an accurate estimate is obtained by solving only one linear program.

An additional difficulty of this sequence of images is that there are some "wrong" scene points which have small reprojection error. It is noteworthy that the number of this

---

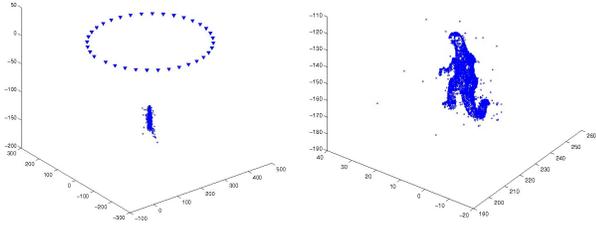[4]http://www.robots.ox.ac.uk/~vgg/data1.html

Figure 2. *Left: the positions of cameras and the scene points estimated by our method. Right: zoom on the scene points.*

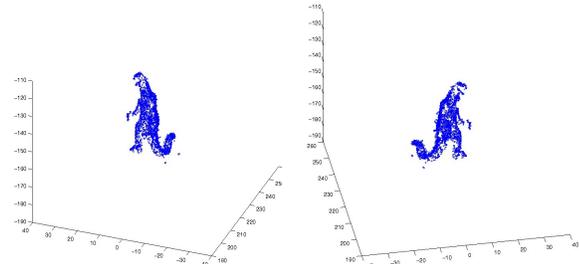Figure 3. *The scene points estimated by our method after removing the points seen by only two cameras.*

Figure 4. *Upper view of the scene points estimated by the method from [15] (left panel), by our method (central panel) and after removing the points seen by only two cameras (right panel).*

Figure 5. *Boxplots of the errors of the estimated camera locations. Right: our procedure vs KK-procedure [15]. Left: our procedure vs SH-procedure.*

kind of outliers can be drastically reduced by considering only those 3D scene points for which we have at least three 2D image points. For the dinosaur sequence, this reduces the number of scene points from 4.983 to 2.307. These points are plotted in Fig. 3, which shows that there is no flagrant outlier in this reduced data set.

For comparison, we ran on the same data the procedures proposed by [24, 15] that we will refer to as SH-procedure and KK-procedure, respectively. To remove 1,500 outliers, the SH-procedure required more than 50 cycles, each cycles consisting of a bisection algorithm containing between 5 and 10 LPs. The resulting estimator had a maximal reprojection error equal to 1.33 pixel. The boxplots of the errors for estimated camera locations are shown in Fig. 5. Concerning the KK-procedure [15], it produces an estimator which is nearly as accurate as the one presented in this paper, but requires a good estimate of the number N_O of outliers, does not contain a step of outlier removal and needs solving more LPs to find the estimator. In this example, we ran the KK-procedure with $m = N - N_O = 15.000$, which is approximately the number of inliers detected by our method. The results presented in Fig. 4 and 5 show that our procedure compares favorably to that of [15].

This experiment clearly demonstrated the competitivity of the proposed methodology with the approaches proposed in [24] and [15] both in terms of the accuracy and the execution time. Similar behavior have been observed on other datasets, see Fig. 6-7.
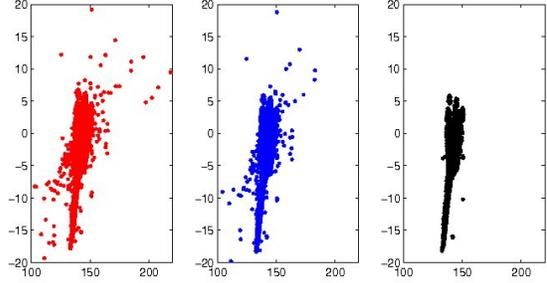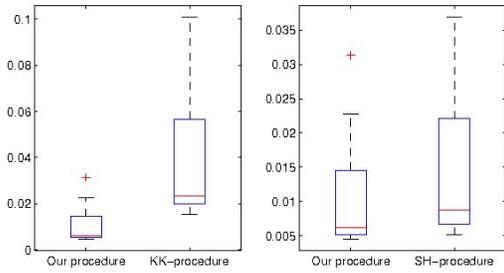
7. Discussion

In this paper, we have shown that the methodology developped for learning sparse representations can be successfully applied to the estimation problems of multiview geometry that are affected by outliers. A rigorous Bayesian framework for the problem of translation estimation and triangulation have been proposed, that have leaded to a new robust estimation procedure. The proposed estimator exploits the sparse nature of the vector of outliers through L_1-norm minimization. We have given the mathematical proof of the result demonstrating the efficiency of the proposed estimator under some assumptions. The relaxation of these assumptions is an interesting theoretical problem that is the object of an ongoing work. Real data analysis conducted on the dinosaur sequence supports our theoretical results and show that our procedure is competitive with the most recent robust estimation procedures.

References

[1] P. Bickel, Y. Ritov, and A. B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector, 2008. *Annals of Statistics*, to appear.

[2] E. Candès and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.*,
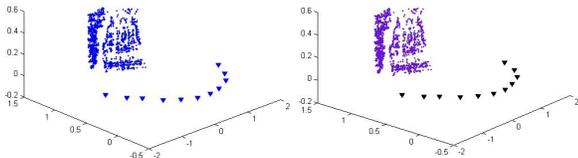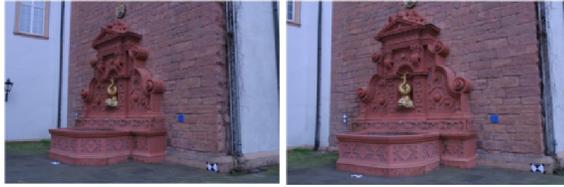
Figure 6. *The Fountain-P11 database [25]. Top: two images out of 11. Bottom left: estimated (by our method) cameras and scene points. Bottom right: true cameras and estimated scene points.*
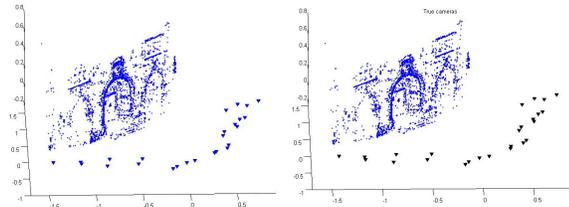

Figure 7. *The Herz-Jesu-P25 database [25]. Top: two images out of 25. Bottom left: estimated (by our method) cameras and scene points. Bottom right: true cameras and estimated scene points.*

35(6):2313–2351, 2007.

[3] E. J. Candès and P. A. Randall. Highly robust error correction by convex programming. *IEEE Trans. Inform. Theory*, 54(7):2829–2840, 2008.

[4] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[5] A. Dalalan and A. Tsybakov. Sparse regression learning by aggregation and langevin monte-carlo. *22th Annual Conference on Learning Theory*, 2009.

[6] A. S. Dalalyan, A. Juditsky, and V. Spokoiny. A new algorithm for estimating the effective dimension-reduction subspace. *Journal of Machine Learning Research*, 9:1647–1678, Aug. 2008.

[7] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1):6–18, 2006.

[8] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, 2001.

[9] O. Enqvist and F. Kahl. Robust optimal pose estimation. In *ECCV*, pages I: 141–153, 2008.

[10] R. Hartley and F. Kahl. Global optimization through rotation space search. *IJCV*, 2009.

[11] R. I. Hartley and F. Schaffalitzky. L_∞ minimization in geometric reconstruction problems. In *CVPR (1)*, pages 504–509, 2004.

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

[13] F. Kahl. Multiple view geometry and the L_∞-norm. In *ICCV*, pages 1002–1009. IEEE Computer Society, 2005.

[14] F. Kahl and R. I. Hartley. Multiple-view geometry under the L_∞ norm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(9):1603–1617, sep 2008.

[15] T. Kanade and Q. Ke. Quasiconvex optimization for robust geometric reconstruction. In *ICCV*, pages II: 986–993, 2005.

[16] Q. Ke and T. Kanade. Uncertainty models in quasiconvex optimization for geometric reconstruction. In *CVPR*, pages I: 1199–1205, 2006.

[17] J. Langford, L. Li, and T. Zhang. Sparse online learning via truncated gradient. *Journal of Machine Learning Research*, 10:777–801, 2009.

[18] H. D. Li. A practical algorithm for L_∞ triangulation with outliers. In *CVPR*, pages 1–8, 2007.

[19] D. Martinec and T. Pajdla. Robust rotation and translation estimation in multiview reconstruction. In *CVPR*, pages 1–8, 2007.

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

[21] C. Olsson, A. P. Eriksson, and F. Kahl. Efficient optimization for L_∞ problems using pseudoconvexity. In *ICCV*, pages 1–8, 2007.

[22] Y. D. Seo and R. I. Hartley. A fast method to minimize L_∞ error norm for geometric vision problems. In *ICCV*, pages 1–8, 2007.

[23] Y. D. Seo, H. J. Lee, and S. W. Lee. Sparse structures in L-infinity norm minimization for structure and motion reconstruction. In *ECCV*, pages I: 780–793, 2008.

[24] K. Sim and R. Hartley. Removing outliers using the L_∞ norm. In *CVPR*, pages I: 485–494, 2006.

[25] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. pages 1–8, 2009.

[26] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999.