Statistical Pose Averaging with Non-isotropic and Incomplete Relative Measurements

Roberto Tron and Kostas Daniilidis

GRASP Lab, University of Pennsylvania, Philadelphia, PA, USA
{tron,kostas}@cis.upenn.edu

Abstract. In the last few years there has been a growing interest in optimization methods for averaging pose measurements between a set of cameras or objects (obtained, for instance, using epipolar geometry or pose estimation). Alas, existing approaches do not take into consideration that measurements might have different uncertainties (i.e., the noise might not be isotropically distributed), or that they might be incomplete (e.g., they might be known only up to a rotation around a fixed axis). We propose a Riemannian optimization framework which addresses these cases by using covariance matrices, and test it on synthetic and real data.

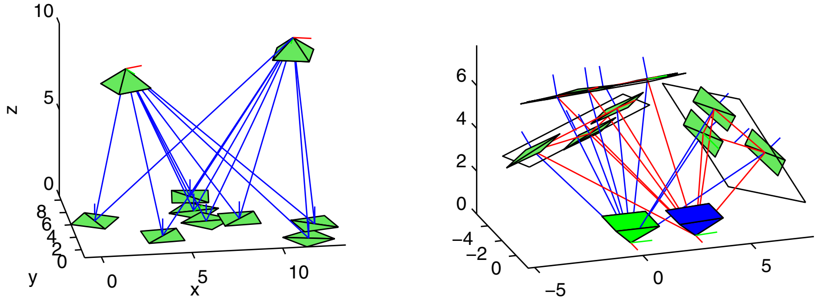
Keywords: Pose averaging, Riemannian geometry, Error propagation, Anisotropic filtering, Incomplete measurements.

1 Introduction

Consider N reference frames, each representing, e.g., the pose of a camera or of an object. Assume that we can completely or partially measure the relative rigid body transformations for a subset of all possible pairs of frames (see Figure 1). Our goal is to combine all these measurements and obtain an estimate of the position of each frame with respect to some global reference. In order to do so, if there are enough measurements available, we can exploit the geometric constraints induced by combining the poses in cycles. This usually takes the form of an optimization problem that "averages" the poses. However, we need to take into account that the estimates might be partially erroneous or unknown. For instance, the noise in the estimated translations could be higher in some direction, or two rotations could be constrained to be coplanar and have the same z-axis, but differ otherwise. If these errors and ambiguities are not correctly handled, they could propagate and bias the entire result. However, if correctly combined, the different measurements can complement each other into a complete and accurate solution. In this paper we propose to explicitly model non-isotropic noise and incomplete poses through the use of covariance matrices. This is similar to the idea of *gradient-weighted least-squares fitting* in the statistics literature [22]. More in detail, we propose to proceed as follows:

1. Estimate the relative rigid body transformations between pairs of references and their uncertainties or ambiguities.

D. Fleet et al. (Eds.): ECCV 2014, Part V, LNCS 8693, pp. 804–819, 2014.



(a) Camera-Targets (CT) dataset.

(b) Camera-Depth sensor-Target (CDT) dataset.

Fig. 1. Example of different graphs of relative poses. Green pyramids: cameras. Green squares: targets with known structure. Blue pyramids: depth sensors. White squares: planes. Blue lines: complete pose estimates. Red lines: incomplete pose estimates. More details on this examples are given in §7.1.

2. Setup a graphical model for the joint probability of the poses.
3. Optimize the cost function associated with this model.

Note that our unknowns are poses, which lie on a Riemannian manifold. We will show in this work how this affects the steps above.

Prior Work. The idea of estimating covariance matrices to obtain better estimation accuracies has been a long-standing idea, even in the context of computer vision [23], [7]. In this work we use similar ideas, although from the more formal setting of Riemannian geometry.

Regarding pose averaging, this problem has been considered in numerous existing papers. The first contributions tried to solve the problem linearly using quaternions [11] or rotation matrices [15]. Following works are based on local optimization approaches: [12] uses Lie groups approximations, [19] applies gradient descent on the sum of squared Riemannian distances (i.e., an ℓ_2 norm), while [13, 20] use absolute distances (i.e., an ℓ_1 norm). In practice, the latter approaches are sensitive to local minimizers. Therefore, more recent work tackles the problem of obtaining good initializations: [4] solves a discretization of the problem using Belief Propagation, while [8] uses a formulation based on Lagrangian duality and [2] proposes a spectral method and an SDP relaxation. Closer to our approach is [5], where Belief Propagation is used for inference on the graphical models of the poses. However, no conditions for a consistent solution were imposed and the non-Euclidean structure of the rotations was not rigorously considered. Finally, Cramer-Rao bounds for this type of problems are considered in [3].

Paper Contributions. In almost all the previous work, the authors make the implicit or explicit assumptions that either the noise in the measurements is isotropically distributed or that the measurements themselves are complete. The major contribution of the present paper is to propose a feasible way to remove these assumptions. In addition, we give in §5 detailed formulae for the estimation of the covariance matrices from image data (while taking into account the Riemannian structure of the space of poses), and, in §6.2, a linear solution to the pose estimation problem from incomplete measurements (plane/pose correspondences).

2 Notation

We model the set of frames and measurements as a directed graph $G = \{V, E\}$, where the vertices $V = \{1, \dots, N\}$ represent the reference frames and the edges $E \subseteq V \times V$ represent the pairs of nodes for which we have a rigid body transformation measurement. We denote the pose of each reference frame as a pair $g_i = (R_i, T_i) \in SE(3)$, where $SE(3) = SO(3) \times \mathbb{R}^3$ is the space of 3-D poses and $SO(3) = \{R \in \mathbb{R}^{3\times 3} : R^T R = I, \det(R) = 1\}$ is the space of 3-D rotations. We also denote as $\tilde{g}_{ij} = (\tilde{R}_{ij}, \tilde{T}_{ij}) \in SE(3)$, $(i,j) \in E$, the measured relative rigid body transformations. We use the convention that (R_i, T_i) represents the transformation from local to global reference frames. Then, ideal, complete and noiseless measurements are given as $g_{ij} = \big(R_{ij}, T_{ij}\big) = g_i^{-1} g_j = \big(R_i^T R_j, R_i^T(T_j - T_i)\big)$. The *tangent space* at $g \in SE(3)$ is given by $T_g SE(3) = T_R SO(3) \times \mathbb{R}^3$, where $T_R SO(3) = \{RV : V \in \mathfrak{so}(3)\}$ is the tangent space at $R \in SO(3)$ and where $\mathfrak{so}(3)$ is the space of 3×3, skew symmetric matrices. We can identify a tangent vector $V_R \in T_R SO(3)$ with a vector $v_R \in \mathbb{R}^3$ using the *hat* $(\cdot)^\wedge$ and *vee* $(\cdot)^\vee$ operators, given by the relations

$$v_R = \begin{bmatrix} v_{R1} \\ v_{R2} \\ v_{R3} \end{bmatrix} \underset{(\cdot)^\vee}{\overset{(\cdot)^\wedge}{\rightleftarrows}} V_R = R \begin{bmatrix} 0 & -v_{R3} & v_{R2} \\ v_{R3} & 0 & -v_{R1} \\ -v_{R2} & v_{R1} & 0 \end{bmatrix}. \tag{1}$$

Similarly, we can identify a vector $V = (\hat{v}_R, v_T) \in T_g SE(3)$ with a vector $v = \operatorname{stack}(v_R, v_T) \in \mathbb{R}^6$. Given two tangent vectors $V_1, V_2 \in T_g SE(3)$ and the corresponding vector representations $v_1, v_2 \in \mathbb{R}^6$, we use the following Riemannian metric for $SE(3)$:

$$\langle V_1, V_2 \rangle = v_1^T v_2. \tag{2}$$

This metric is equivalent to consider $SE(3)$ as the cartesian (instead of semi-direct) product of $SO(3)$ and $\mathbb{R}^3$, and then summing the standard metrics of the two.

For a given $R \in SO(3)$, the *exponential* and *logarithm* map are denoted, respectively, as $\exp_R : T_R SO(3) \to SO(3)$ and $\log_R : U_R \to T_R SO(3)$, where $U_R \subset SO(3)$ is the maximal set containing R for which $\exp_R$ is diffeomorphic. For convenience, we also define $\operatorname{Log} : U_I \to \mathbb{R}^3$, the vectorized version of the logarithm map at the identity, i.e., $\operatorname{Log}(R) = (\log_I(R))^\vee \in \mathbb{R}^3$, where $R \in SO(3)$. For any given rotation $R \in SO(3)$, we denote as $\operatorname{DLog}(R)$ the matrix representation of the differential of the logarithm. More precisely, let $R(t)$ be a smooth curve such that $R(0) = R_0$ and $\dot{R}(0) = \hat{v}_R$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} \operatorname{Log}(R)|_{t=0} = \operatorname{DLog}(R_0) v_R. \tag{3}$$

The operators Log and DLog can be computed in closed-form [18].

For a given $g \in SE(3)$, the exponential and logarithm map are defined by the same maps applied independently on the rotation and translation components:

$$\exp_{(R,T)}(\hat{v}_R, v_T) = (\exp_R \hat{v}_R, T + v_T), \tag{4}$$

$$\log_{(R_1,T_1)}(R_2, T_2) = (\log_{R_1} R_2, T_2 - T_1). \tag{5}$$

Note that these definition are consistent with the choice of Riemannian metric in (2), and are different than the so-called *screw-motions* (see, e.g., [14]). If $\tilde{g} = (\tilde{R}, \tilde{T}) \in SE(3)$ is a random variable with distribution $p(\tilde{g})$, the covariance matrix with respect to $g = (R, T)$ (which is conventionally taken to be the mean) is defined as

$$\Sigma = \int_{SE(3)} v_{\tilde{g}g} v_{\tilde{g}g}^T p(\tilde{g}) \mathrm{d}SE(3), \tag{6}$$

where $v = \operatorname{stack}(\operatorname{Log}(\bar{R}^T \tilde{R}), \tilde{T} - \bar{T})$ and $\mathrm{d}SE(3)$ denotes the measure induced by the Riemannian metric [17]. Note that v is not defined on the so-called *cut locus* of g (i.e., outside of the set U_R mentioned above). However, it can be shown [9] that the cut locus has measure zero. Hence, for well-behaved distributions (such as those considered here), the integral in (6) is well defined.

Given a function $f : SE(3) \to \mathbb{R}^D$, we denote as $\operatorname{grad}_g f$ a unique set of tangent vectors $\{\operatorname{grad}_g^d f\}_{d=1}^D$ such that, for any tangent vector $v \in T_g SE(3)$, we have

$$\langle \operatorname{grad}_g^d f, v \rangle = \frac{\mathrm{d}}{\mathrm{d}t} f_d(\tilde{g}(t)) \bigg|_{t=0}, \tag{7}$$

where f_d is the d-th component of f and $\tilde{g}(t)$ is a curve in $SE(3)$ such that $\tilde{g}(0) = g$ and $\dot{\tilde{g}}^\vee = v$. In practice, using the identification given by the hat operator, $\operatorname{grad} f$ can be expressed as a $6 \times d$ matrix. This is equivalent to the definition of gradient and Jacobian for the case where f is defined on the usual Euclidean space.

Lastly, we use I_d and 0_d to denote the identity and square zero matrix in $\mathbb{R}^{d \times d}$, respectively. Also, we denote with $\{e_d\}_{d=1}^3$ and $\{e'_d\}_{d=1}^2$ the standard bases in $\mathbb{R}^3$ and $\mathbb{R}^2$.

3 Problem Setting

For complete measurements $\tilde{g}_{ij}$, we assume a clipped Gaussian distribution with mean at the true values g_{ij}. This distribution has the form

$$p(\tilde{g}_{ij}; g_i, g_j) = k_{ij} \exp\left(-\frac{v_{ij}^T \Gamma_{ij} v_{ij}}{2}\right) \tag{8}$$

where

$$v_{ij} = \begin{bmatrix} v_{Rij} \\ v_{Tij} \end{bmatrix}, \quad v_{Rij} = \operatorname{Log}\big(\tilde{R}_{ij}^T R_i^T R_j\big), \quad v_{Tij} = \tilde{T}_{ij} - R_i^T (T_j - T_i), \tag{9}$$

Notation-wise, we will use the following partition of Γ_{ij}:

$$\Gamma_{ij} = \begin{bmatrix} \Gamma_{RRij} & \Gamma_{RTij} \\ \Gamma_{RTij}^T & \Gamma_{TTij} \end{bmatrix}, \tag{10}$$

where $\Gamma_{RRij}, \Gamma_{RTij}, \Gamma_{TTij} \in \mathbb{R}^{3\times 3}$. The clipped Gaussian has been shown [17] to be a generalization of the usual Gaussian distribution in $\mathbb{R}^D$ to the manifold case, in the sense that it is the distribution that maximizes the entropy for a given covariance matrix Σ_{ij}. The constant k_{ij} is chosen such that $p(g_{ij})$ integrates to one over $SE(3)$ [17]. If the distribution is relatively concentrated (intuitively, far from a uniform distribution), the *dispersion* or *information matrix* Γ_{ij} is related, as a first order approximation [17], to the covariance matrix by the formula

$$\Gamma_{ij} \simeq \Sigma_{ij}^{-1} - \frac{1}{3}\,\mathrm{Ric}, \tag{11}$$

where Ric is the matrix form in normal coordinates of the *Ricci* (or *scalar*) *curvature tensor*. In our case, we have the following:

Proposition 1. *With the choice of metric for $SE(3)$ made in §2, we have*

$$\mathrm{Ric} = \begin{bmatrix} \frac{1}{2}I_3 & 0_3 \\ 0_3 & 0_3 \end{bmatrix}. \tag{12}$$

See the additional material for a proof. Intuitively, the correction in (11) takes into account the effect of clipping the tails of the Gaussian distribution.

For incomplete measurements, we model the unknown degrees of freedom as directions along which the covariance Σ_{ij} is infinite. In practice, this leads to dispersion matrices Γ_{ij} which are singular along the same directions.

Under the additional assumption that the noise terms affecting different edges $(i,j) \in E$ are independent, the joint distribution of the measurements is given by

$$p(\{\tilde{g}_{ij}\}_{(i,j)\in E}; \{g_i\}_{i\in V}) = \prod_{(i,j)\in E} p(\tilde{g}_{ij}; g_i, g_j). \tag{13}$$

We then formulate the problem of estimating the poses $\{g_i\}_{i\in V}$ as a maximum likelihood problem, i.e., as minimizing the following log-likelihood.

$$l(\{g_i\}_{i\in V}) = -\frac{1}{2} \sum_{(i,j)\in E} v_{ij}^T \Gamma_{ij} v_{ij}. \tag{14}$$

Note that in (14) we have excluded the constant terms containing the normalization constants k_{ij}, because they are not relevant for the optimization problem.

4 Minimization Algorithm

For the minimization of the cost (14), we will employ Riemmannian gradient descent with exact line search. The general form of this algorithm is given in

Algorithm 1. Riemannian gradient descent with exact line search

```
Input: Initial elements $\{g_{0i}\}_{i\in V} \in SE(3)^N$, a maximum step size $\bar{t}$.

1. Initialize $g_i(0) = g_{0i}$
2. For $l \in \mathbb{N}$, repeat
   (a) Compute the gradient $\{h_i\}_{i\in V}$ defined by (15)
   (b) Solve the line search problem
       $t^* = \operatorname{argmin}_{t\in[0,\bar{t}]} l(\{\exp_{g_i(l)}(th_i)\}_{i\in V}$    (19)
   (c) Compute the update
       $g_i(l+1) = \exp_{g_i(l)}(t^* h_i), \quad i \in V$    (20)
```

Algorithm 1. Other variations are also possible (e.g., using an inexact line search or a fixed step size). See [1] for details on such variations and for convergence proofs. The only part of the algorithm that is specific to our problem is in the computation of the gradient, which is given by the following.

Proposition 2. *The gradient of the negative log-likelihood function* (14) *is given by*

$$\operatorname{grad}_{g_i} l(\{g_i\}_{i\in V}) = \\ \Bigg(\sum_{(i,j)\in E} (R_i^T R_j h'_{Rij} + h''_{Rij}) - \sum_{(j,i)\in E} h'_{Rji}, \sum_{(j,i)\in E} h_{Tji} - \sum_{(i,j)\in E} h_{Tij}\Bigg) \tag{15}$$

where

$$h'_{Rij} = \mathrm{DLog}(R_{ij}^T R_i^T R_j)^T (\Gamma_{RRij} v_{Rij} + \Gamma_{RTij} v_{Tij}) \tag{16}$$

$$h''_{Rij} = \big(Ri^T(T_j - T_i)\big)^\wedge (\Gamma_{RTij}^T v_{Rij} + \Gamma_{TTij} v_{Tij}) \tag{17}$$

$$h_{Tij} = R_i(\Gamma_{RTij}^T v_{Rij} + \Gamma_{TTij} v_{Tij}); \tag{18}$$

See the Appendix for a proof.

4.1 Numerical Normalization

Notice that terms in (15) depend directly on the entries of $\{\Gamma_{ij}\}_{(i,j)\in E}$, which can be relatively large or small when the measurements are either really precise (small variances) or not precise at all (large variances). In practice, this has the effect that the step size t found by Algorithm 1 could become very small or very large. In order to avoid potential numerical problems that could arise, we rescale the cost (14) by performing the substitution

$$\Gamma_{ij} \leftarrow \frac{\Gamma_{ij}}{\mu_\Gamma} \tag{21}$$

where

$$\mu_\varGamma = \frac{1}{6|E|} \sum_{(i,j)\in E} \operatorname{tr}(\varGamma_{ij}). \tag{22}$$

With this normalization the average of the eigenvalues of $\varGamma_{ij}$ will be equal to one. Note that this rescaling does not affect where the minimizers of (14) are located.

5 Computing the Dispersion Matrices

In this section, we give details on how to compute the dispersion matrices $\varGamma_{ij}$. We consider the cases where we estimate complete poses from image data and where we have incomplete pose constraints.

5.1 Complete Poses from Image Data

In most applications, each relative transformation $\tilde{g}_{ij}$ is estimated as the solution to some minimization problem of the form

$$\tilde{g}_{ij} = \operatorname*{argmin}_{g_{ij}} \sum_{k=1}^{K} f(g_{ij}, x_{ij}^{(k)}), \tag{23}$$

where $\{x_{ij}^{(k)}\}_{k=1}^{K}$ represents a set of K measurements in $\mathbb{R}^D$ (e.g., matched image point coordinates). Then, one can think of the noise in $\tilde{g}_{ij}$ as being a consequence of the noise present in the data $x_{ij}^{(k)}$. We can extend well-known error propagation techniques (see, e.g., [7, 22, 23]) to the case of manifolds. Define $\nabla f(g_{ij}, x_{ij}^{(k)}) = \operatorname{grad}_{g_{ij}} f(g_{ij}, x_{ij}^{(k)})$. The solution $\tilde{g}_{ij}$ of (23) implies the condition

$$\sum_{k} \nabla f(g_{ij}, x_{ij}^{(k)}) = 0 \tag{24}$$

Writing the first order Taylor expansion of (24) with respect to both g_{ij} and $x_{ij}^{(k)}$, we have

$$\sum_{k=1}^{K} H_{ijk} v_{ij} + \sum_{k=1}^{K} J_{ijk} v_{xij}^{k} \simeq 0, \tag{25}$$

where

$$H_{ijk} = \operatorname{grad}_{g_{ij}} \nabla f(g_{ij}, x_{ij}^{(k)}) \tag{26}$$

$$J_{ijk} = \operatorname{grad}_{x_{ij}^{(k)}} \nabla f(g_{ij}, x_{ij}^{(k)}) \tag{27}$$

and $v_{ij} \in T_{g_{ij}}SE(3)$, $v_{xij}^{k} \in \mathbb{R}^D$ represent the errors in the estimate g_{ij} and the measurements $x_{ij}^{(k)}$, respectively. Under the assumption that errors in the image point are corrupted by an isotropic Gaussian noise with zero mean and covariance $E[v_{xij}^k v_{xij}^{k\,T}] = \sigma I_4$ for some $\sigma \in \mathbb{R}$, we can compute the covariance matrix of v_{ij} as

$$\varSigma_{ij} = E[v_{ij} v_{ij}^T] = (\sum_{k=1}^{K} H_{ijk})^{-1} (\sum_{k=1}^{K} J_{ijk} J_{ijk}^T) (\sum_{k=1}^{K} H_{ijk})^{-T}. \tag{28}$$

This result can then be plugged into (11) to obtain the dispersion matrices for the optimization problem.

As a concrete example, consider the estimation of the pose of a camera i from an object j with known geometry by minimizing the sum of reprojection errors

$$f(R_{ij}, T_{ij}, x_{ij}^{(k)}) = \|\pi\big(R_{ij}^T (X_{ij}^{(k)} - T_{ij})\big) - x_{ij}^{(k)}\|^2, \tag{29}$$

where X is the 3-D coordinate vector of a known point in the object's reference frame, x is its (measured) projection in the camera's image and π represents the perspective projection operator with unit focal length. The camera is assumed to be calibrated. Then, we have the following.

Proposition 3. *The matrices J_{ijk} and H_{ijk} in* (28) *for the cost* (29) *are*

$$H_{ijk} = J_x^T J_x + \sum_{d=1}^{2} (x_p - x_{ij}^{(k)})^T e_d H_{xd}, \qquad J_{ijk} = -J_x, \tag{30}$$

where

$$J_x = \frac{1}{\lambda_c^2} P M J_X w, \qquad \lambda_c = e_3^T X_c \tag{31}$$

$$H_{xd} = \frac{1}{\lambda_c^2} \left(\frac{1}{\lambda_c} J_X^T e_3 e_d^T M J_X + J_X^T (e_3 e_d^T - e_d e_3^T) J_X + \sum_{d'=1}^{3} (e_d^T M e_{d'} H_{Xd'}) \right), \tag{32}$$

$$M = (\lambda_c I_3 - X_c e_3^T), \qquad P = \begin{bmatrix} I_2 \; 0_2 \end{bmatrix}, \qquad x_p = \frac{1}{\lambda_c} P X_c \tag{33}$$

$$J_X = \begin{bmatrix} \hat{X}_c \; R_{ij}^T \end{bmatrix}, \qquad H_{Xd} = \begin{bmatrix} \hat{X}_c \hat{e}_d & \hat{e}_d^T R_{ij}^T \\ R_{ij} \hat{e}_d & 0_3 \end{bmatrix}, \quad X_c = R_{ij}^T (X_{ij}^{(k)} - T_{ij}). \tag{34}$$

See the additional material for a proof. Note that this result could also be used in PnP problems or in a bundle-adjustment iteration when the 3-D structure is fixed.

5.2 Partial Pose Constraints

In different applications, especially those combining sensors of different modalities, we can only obtain partial constraints between pairs of poses. For instance, consider the problem of calibrating the extrinsic transformation between a depth sensor and a camera using measurements from a plane in different positions (see Figure 1b, for instance). Using a known pattern on the plane, one can compute the relative transformation between the camera frame and a frame attached to the pattern. From the depth sensor, however, we can only estimate the equation of the plane. The relative transformation between the depth sensor and the plane frames can therefore be obtained only up to an in-plane rotation and translation. Similarly, consider the case where multiple known patterns are attached to the plane. We can then add the constraint that the reference frames should differ only by an in-plane rotation and translation.

Mathematically, these cases can be considered as follows. Assume that the measurement $\tilde{g}_{ij}$ is without noise, but with an ambiguity given by a rotation and translation in a plane with normal n, $\|n\| = 1$. We then have $\tilde{g}_{ij} = g_{ij} g_n$, where $g_n = (R_n(\theta), v_n)$, $R_n(\theta)$ is a rotation around of θ radiants around n and v_n is a vector orthogonal to n, i.e., $v^T n = 0$. Note that $\operatorname{Log}\big(R_n(\theta)\big) = \theta n$. It then follows that the tangent vector v_{ij} is given by

$$v_{ij} = \begin{bmatrix} \theta n \\ v_n \end{bmatrix}. \tag{35}$$

We then propose to use the following dispersion matrix

$$\varGamma_{ij} = \begin{bmatrix} I_3 - nn^T & 0_3 \\ 0_3 & nn^T \end{bmatrix}. \tag{36}$$

It is then easy to check that the term $v_{ij}^T \varGamma_{ij} v_{ij}$ in (14) is zero if and only if v_{ij} is of the form (35), i.e., this quantity can be used to measure the similarity between the poses $g_i^{-1} g_j$ and $\tilde{g}_{ij}$.

6 Initialization

Since Algorithm 1 represents a local optimization, using good initial values for $\{g_i\}_{(i,j)\in E}$ is important. These can be found by using linear algorithms. We consider in this section the two cases of complete or incomplete pose estimates.

6.1 Complete Poses

For this case, we can use the spectral method of [2] to estimate the rotations. For the translations, we solve the least squares problem

$$\min_{\{T_i\}} \sum_{(i,j)\in E} \|\tilde{T}_{ij} - R_i^T (T_j - T_i)\|^2, \tag{37}$$

which can be rewritten in the form

$$\min_T \frac{1}{2} T^T A T + b^T T, \tag{38}$$

where $T = \mathrm{stack}(\{T\}_{i \in V}) \in \mathbb{R}^{3N}$, and where the matrix $A \in \mathbb{R}^{3N \times 3N}$ and the vector $b \in \mathbb{R}^{3N}$ can be computed from $\{R_i\}_{i \in V}$ and $\{T_{ij}\}_{(i,j) \in E}$. Note that A is singular, due to the fact that if $\{T_i\}_{i \in V}$ is a minimizer of (37), then also $\{T_i + \tilde{T}\}_{i \in V}$ will be a minimizer for any $\tilde{T} \in \mathbb{R}^3$. In practice, one needs only to compute $T = A^{\dagger} b$, where $A^{\dagger}$ is the Moore-Penrose pseudoinverse of A.

6.2 Incomplete Poses

Assume that we have a node i for which all the measurements $\{g_{ij}\}_{j:(i,j) \in E}$ contain the ambiguity described in §5.2 and all the poses $\{g_i\}_{j:(i,j) \in E}$ are known (e.g., they can be directly measured). Assume that the equation of the plane for the ambiguity in g_{ij} is described as $n_{ij}^T X_i + d_{ij} = 0$ in the frame g_i, and $n_{ji}^T X_j + d_{ji} = 0$ in the frame g_j. Then, we have

$$n_{ij}^T R_{ij} X_j + n_{ij}^T T_{ij} + d_{ij} = 0, \tag{39}$$

from which we can obtain the following constraints:

$$R_i n_{ij} = R_j n_{ji}, \tag{40}$$

$$n_{ij}^T R_i^T (T_j - T_i) + d_{ij} = d_{ji}. \tag{41}$$

It follows that we can first estimate R_i through (40) and an orthogonal Procrustes analysis [10]. In practice, defining $C = \sum_{j:(i,j) \in E} R_i n_{ij} n_{ji}^T$, we can set $R_i = U \operatorname{diag}\bigl(1, 1, \det(UV^T)\bigr) V^T$, where U, V are given by the SVD decomposition $C = U \Sigma V^T$. Once R_i is determined, we can use (41) to estimate T_i linearly in a least-squares sense.

7 Experiments

7.1 Synthetic Datasets

We first test the minimization algorithm described in §4 on two synthetic datasets.

Cameras-Targets (CT) Dataset. In the first dataset, we generate $N = 10$ reference frames where two frames (which represent "cameras") lie on the $z = 10$ plane and point downward while the remaining eight frames (which represent "targets" with known structure) lie on the plane $z = 0$ and point upward. The x and y coordinates of each frame are taken uniformly at random in a square with side length $L = 15$. The graph is formed by connecting each camera with the other and with each target. An example of this setup is depicted in Figure 1a. This scenario is meant to represent two poses of a camera moving above targets with known 3-D structure but unknown location. We generate each dispersion matrix (for the rotation part) as $\Gamma_{RRij} = U \Lambda^{-1} U^T$, where $U \in \mathbb{R}^{3 \times 3}$ is drawn uniformly at random from the space of orthonormal matrices and $\Lambda \in \mathbb{R}^{3 \times 3}$ is a diagonal matrix whose elements are variances drawn uniformly at random between 0.01 and 0.5 radiants (which are roughly equivalent to 0.5 and 30 degrees, respectively). A similar approach is taken for the translation part Γ_{TTij}, but with variances drawn uniformly between 0.1 and 2. For the cross-covariance matrices Γ_{RTij}, each entry is taken from a Gaussian distribution with standard deviation 0.01. The measurement $(\tilde{R}_{ij}, \tilde{T}_{ij})$ for each edge is generated by corrupting the true transformation according to the model in (8) and the generated covariances $\{\Gamma_{ij}\}_{(i,j)\in E}$. We then run the minimization algorithm of §4 on the generated dataset, first using the true covariance matrices $\{\Gamma_{ij}\}_{(i,j)\in E}$ ("Covariances"), and then again with $\Gamma_{ij} \leftarrow \frac{\mathrm{tr}(\Gamma_{ij})}{6} I_6$ ("Weighted") and $\Gamma_{ij} \leftarrow I_6$ ("Isotropic") for all $(i,j) \in E$. The last two configurations correspond, respectively, to just using the variance information without considering any coupling between rotation and translation, and to assuming that the errors are all isotropic and identically distributed. We repeated this experiment 50 times and, for each edge, we computed the angles between the relative rotations and between the relative translations from the poses obtained with the minimization algorithms (including the spectral initialization of §6.1) and those obtained from the ground truth. Figure 2 shows the empirical distributions of these errors, while Table 1 contains the corresponding mean and standard deviations. As a reference, we also include the same statistics for the initialization described in §6.1 ("Spectral").

From Figures 2a, 2b and Table 1a we can see that by averaging the measurements across the graph the errors (especially for the translations) are greatly reduced, both in mean and standard deviation. Also, by taking into account the correct covariance matrices, we obtain around 40% reduction in the mean error for the rotations and around 20% reduction for the translations. Intuitively, this is because our algorithm uses the fact that different measurements have different uncertainties in different directions.

Camera-Depth sensor-Targets (CDT) dataset. In the second dataset, we simulate a camera-depth sensor extrinsic calibration task. We pose the problem by using a plane with three targets on it and assume that each the camera can independently measure the pose of each target, while the depth sensor can only measure the normal and distance of the plane on which the targets reside. The

Table 1. Mean and standard deviation for the errors after localization and for the initial measurements. The errors are in degrees between ground truth and estimated rotation axes. The same applies to the translation directions.

(a) CT dataset

	Rotations		Translations	
	Mean	Std	Mean	Std
Covariances	4.130	2.041	1.856	1.063
Weighted	6.259	4.061	2.315	1.429
Isotropic	6.369	4.079	2.306	1.421
Spectral	9.208	4.381	3.971	2.267

(b) CDT dataset

	Rotations		Translations	
	Mean	Std	Mean	Std
Covariances	1.019	0.538	1.114	0.686
Linear	1.070	0.491	1.544	1.055

in-plane transformation between the targets is assumed unknown. We generate 20 random plane positions, and corrupt the corresponding measurements with an isotropic noise with variance 0.05 for both rotation and translation in the pose estimates, and both normal and distance in the plane estimates. We build the graph for our algorithm with edges between the camera and each target (complete poses), between each target and the depth sensor (incomplete poses) and between targets belonging to the same plane (incomplete poses). An illustration of this setup (with only three planes visualized) can be found in Figure 1b. We initialize our algorithm by fixing the camera frame at the origin, the targets' frames at their measured poses and the depth sensor frame at the solution given by the linear algorithm of §6.2. We repeat the localization for 50 different plane configurations, and measure the error on the rotation and translation direction for the pose of the depth sensor with respect to the camera. Similarly to before, Figures 2c, 2d and Table 1b show the cumulative distributions and the mean and standard deviation of the errors. Again, our algorithm improves the localization results with respect to the linear initialization, especially for the translation part.

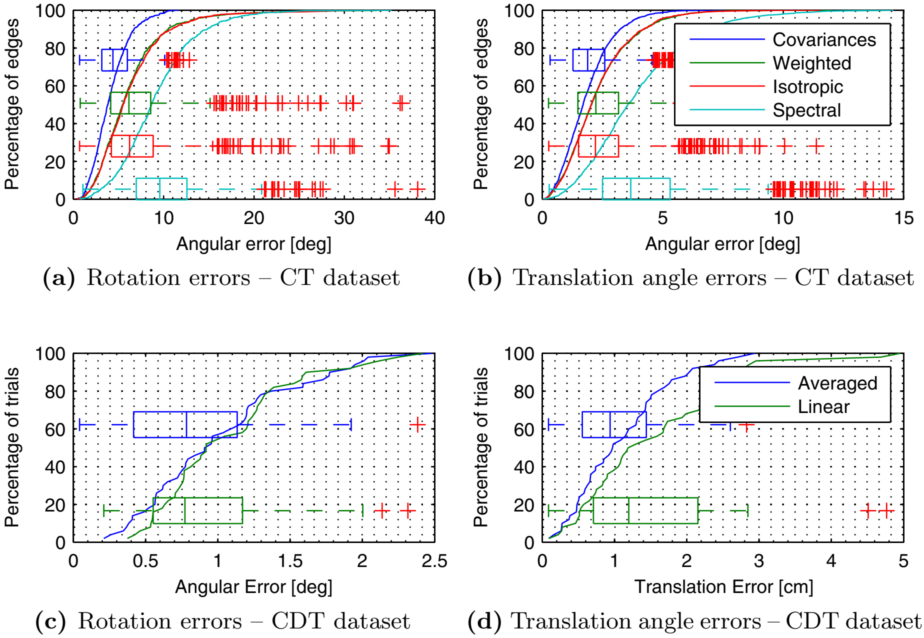
7.2 Real Dataset

The dataset for this section has been obtained by collecting a sequence of 65 images of 6 AprilTags [16] (our targets) from a moving camera strapped onboard a quadcopter. The ground truth pose of each tag has been collected using a commercial VICON system, while the measurements of the relative pose between each visible tag and the camera are obtained using the software of [16] followed by a routine to minimize the reprojection error of the four corners of each tag. This dataset is particularly challenging because we have only four points available for pose estimation for each target and the dataset also contains a few mis-detections, which introduce outliers.

For both datasets, we use the method described in §5 to compute the covariance matrices for each estimate and run the same algorithms as in §7.1. For the evaluation, we consider only the nodes of the graph corresponding to the targets. Since we are interested in the error on the absolute pose, we align the result of each optimization to the ground truth using a Procrustes analysis. We record the root mean square error of the translations and the angle between the rotations. The results are reported in Table 2. the result using full covariance matrices ("Covariances") is strictly better than the result using equally weighted isotropic

Table 2. Errors on the pose of the tags (average angle for rotations, and root mean squared error for translations) with respect to ground truth for the Tag dataset.

	Rotations [deg]	Translations [cm]
Covariances	2.066	2.378
Weighted	3.293	1.853
Isotropic	2.331	2.670
Spectral	1.559	2.563



(a) Rotation errors – CT dataset **(b)** Translation angle errors – CT dataset

(c) Rotation errors – CDT dataset **(d)** Translation angle errors – CDT dataset

Fig. 2. Empirical cumulative distributions and box plots of the errors after localization. The various lines correspond to the estimation with exact, weighted and isotropic covariances and to the spectral initialization.

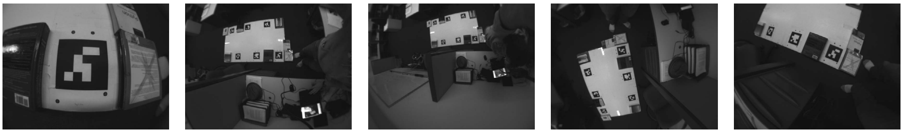


Fig. 3. Some of the images in the Tag dataset.

covariances ("Isotropic"). Using only the trace of the covariances ("Weighted") gives mixed results, with better translations but significantly worse rotations. We believe this might be due to the fact that, from only four points, the estimate of the covariance matrices is less indicative of the true covariance, and using only the trace might give better results. Finally, in both cases the initialization with the algorithm of [2] gives better rotation estimates but worse translation estimates. This might due to the fact that the datasets contain outliers due to wildly incorrect pose estimations, and the Frobenious-norm distance used for the rotations in [2] is slightly more robust to outliers than the Riemannian distance used here.

8 Conclusions and Future Work

In this paper we proposed an optimization method for averaging poses along a graph with non-istropic and incomplete measurements. Our technique can be seen as the extension of weighted least squares to this problem. We have also examined a method for analytically estimating the necessary covariance matrices in the case of pose estimation from objects with known geometry or with partial plane-pose constraints. We plan to expand this work by considering noise models with outliers and other kinds of ambiguities in the measurements, possibly combining it with other recent approaches such as [7,23].

Acknowledgements. We gratefully acknowledge the support of grants ARL MAST CTA W911NF-08-2-0004, ARL Robotics CTA W911NF-10-2-0016, NSF-DGE-0966142, and NSF-IIS-1317788.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton(2008)
2. Arie-Nachimson, M., Kovalsky, S., Kemelmacher-Shlizerman, I., Singer, A., Basri, R.: Global motion estimation from point matches. In: International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, pp. 81–88 (2012)
3. Boumal, N.: On intrinsic Cramer-Rao bounds for Riemannian submanifolds and quotient manifolds. IEEE Transactions on Signal Processing 61(7), 1809–1821 (2013)
4. Crandall, D., Owens, A., Snavely, N., Huttenlocher, D.P.: Discrete-continuous optimization for large-scale structure from motion. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3001–3008 (2011)
5. Devarajan, D., Cheng, Z., Radke, R.: Calibrating distributed camera networks. Proceedings of the IEEE 96(10), 1625–1639 (2008)
6. Enqvist, O., Kahl, F., Olsson, C.: Non-sequential structure from motion. In: Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, pp. 264–271 (2011)
7. Faugeras, O.: Three dimensional computer vision: A geometric viewpoint. The MIT Press (1993)
8. Fredriksson, J., Olsson, C.: Simultaneous multiple rotation averaging using lagrangian duality. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part III. LNCS, vol. 7726, pp. 245–258. Springer, Heidelberg (2013)
9. Gallier, J.: Notes on Differential Geometry and Lie Groups (in preparation, 2014), http://www.cis.upenn.edu/%7Ejean/gbooks/manif.html
10. Golub, G.H., Loan, C.F.V.: Matrix computations, vol. 3. Johns Hopkins University Press (1996)
11. Govindu, V.M.: Combining two-view constraints for motion estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 218–225 (2001)
12. Govindu, V.M.: Lie-algebraic averaging for globally consistent motion estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 684–691 (2004)

13. Hartley, R., Aftab, K., Trumpf, J.: L1 rotation averaging using the Weiszfeld algorithm. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
14. Ma, Y.: An invitation to 3-D vision: from images to geometric models. Springer (2004)
15. Martinec, D., Pajdla, T.: Robust rotation and translation estimation in multiview reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
16. Olson, E.: AprilTag: A robust and flexible visual fiducial system. In: IEEE International Conference on Robotics and Automation, pp. 3400–3407. IEEE (May 2011)
17. Pennec, X.: Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. Journal of Mathematical Imaging and Vision 25, 127–154 (2006)
18. Tron, R.: Distributed optimization on manifolds for consensus algorithms and camera network localization. Ph.D. thesis, The Johns Hopkins University (2012)
19. Tron, R., Vidal, R.: Distributed image-based 3-D localization in camera sensor networks. In: IEEE International Conference on Decision and Control (2009)
20. Dai, Y., Trumpf, J., Li, H., Barnes, N., Hartley, R.: Rotation averaging with application to camera-rig calibration. In: Zha, H., Taniguchi, R.-i., Maybank, S. (eds.) ACCV 2009, Part II. LNCS, vol. 5995, pp. 335–346. Springer, Heidelberg (2010)
21. Zach, C., Klopschitz, M., Pollefeys, M.: Disambiguating visual relations using loop constraints. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1426–1433 (2010)
22. Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. Image and Vision Computing 15, 59–76 (1997)
23. Zhang, Z., Faugeras, O.: 3D dynamic scene analysis: a stereo based approach. Springer (1992)

Appendix

This appendix contains the proof of Proposition 2. The proofs for the other results can be found in the additional material. In general, we will need to compute the gradient of some function f defined on $SE(3)$. The approach we use is as follows. First, we define the curves $(R_i(t), T_i(t))$ such that $\frac{\mathrm{d}}{\mathrm{d}t}\big(R_i(t), T_i(t)\big)\Big|_{t=0} = \big(\dot{R}_i(0), \dot{T}_i(0)\big) = (v_{\hat{Ri}}, v_{Ti}) = v$. Then, we compute the directional derivative at $t = 0$ of the function along such curves, which we will denote as $\dot{f}(v)$ (where v indicates the direction of the derivative). The result for the gradient will then follow from the definition in (7). Note that, in order to avoid clutter, we will omit, in the following, the explicit dependency of R_i, T_i, etc., on the variable t.

Proof of Proposition 2

We will need the following derivatives. We will use the fact that $R\hat{v}R^T = (Rv)^\wedge$ for $R \in SO(3)$.

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\tilde{R}_{ij}R_i^T R_j &= R_{ij}\dot{R}_i^T R_j + \tilde{R}_{ij}R_i^T \dot{R}_j = \tilde{R}_{ij}\hat{v}_{Ri}^T R_i^T R_j + R_{ij}R_i^T R_j \hat{v}_{Rj} \\ &= \tilde{R}_{ij}R_i^T R_j(-R_j^T R_i \hat{v}_{Ri} R_i^T R_j + \hat{v}_{Rj}) = \tilde{R}_{ij}R_i^T R_j(-R_j^T R_i v_{Ri} + v_{Rj})^\wedge \quad (42)\end{aligned}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} v_{Tij} = -\hat{v}_{Ri}^T R_i^T (T_j - T_i) - R_i^T (v_{Tj} - v_{Ti}) = \big(R_i^T (T_i - T_j)\big)^\wedge v_{Ri} + R_i^T (v_{Ti} - v_{Tj}) \tag{43}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} v_{Rij} = \mathrm{DLog}(\tilde{R}_{ij} R_i^T R_j)(-R_j^T R_i v_i + v_j) \tag{44}$$

We can now take the derivative of the negative log-likelihood.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} l(\{g_i\}_{i \in V}) &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2} \sum_{(i,j) \in E} (v_{Rij}^T \mathit{\Gamma}_{RRij} v_{Rij} + 2 v_{Rij}^T \mathit{\Gamma}_{RTij} v_{Tij} + v_{Tij}^T \mathit{\Gamma}_{TTij} v_{Tij}) \\ &= \sum_{(i,j) \in E} \Big((v_{Rij}^T \mathit{\Gamma}_{RRij} + v_{Tij}^T \mathit{\Gamma}_{RTij}^T) D_{ij} (-R_j^T R_i v_{Ri} + v_{Rj}) \\ &\qquad - (v_{Rij}^T \mathit{\Gamma}_{RTij} + v_{Tij}^T \mathit{\Gamma}_{TTij}) \big(R_i^T (T_j - T_i)\big)^\wedge v_{Ri} \\ &\qquad\qquad - (v_{Rij}^T \mathit{\Gamma}_{RTij} + v_{Tij}^T \mathit{\Gamma}_{TTij}) R_i^T (v_{Tj} - v_{Ti}) \end{aligned} \tag{45}$$

where $D_{ij} = \mathrm{DLog}(R_{ij}^T R_i^T R_j)$. The expressions in Proposition 2 follow from (45) and the definition of gradient.