Structure-from-Motion Based Hand-Eye Calibration Using L_∞ Minimization

Jan Heller[1] Michal Havlena[1] Akihiro Sugimoto[2] Tomas Pajdla[1]

[1]Czech Technical University,
Faculty of Electrical Engineering,
Karlovo náměstí 13, Prague, Czech Republic
{hellej1,havlem1,pajdla}@cmp.felk.cvut.cz

[2]National Institute of Informatics,
2-1-2 Hitotsubashi,
Chiyoda-ku, Tokyo, Japan
sugimoto@nii.ac.jp

Abstract

This paper presents a novel method for so-called hand-eye calibration. Using a calibration target is not possible for many applications of hand-eye calibration. In such situations Structure-from-Motion approach of hand-eye calibration is commonly used to recover the camera poses up to scaling. The presented method takes advantage of recent results in the L_∞-norm optimization using Second-Order Cone Programming (SOCP) to recover the correct scale. Further, the correctly scaled displacement of the hand-eye transformation is recovered solely from the image correspondences and robot measurements, and is guaranteed to be globally optimal with respect to the L_∞-norm. The method is experimentally validated using both synthetic and real world datasets.

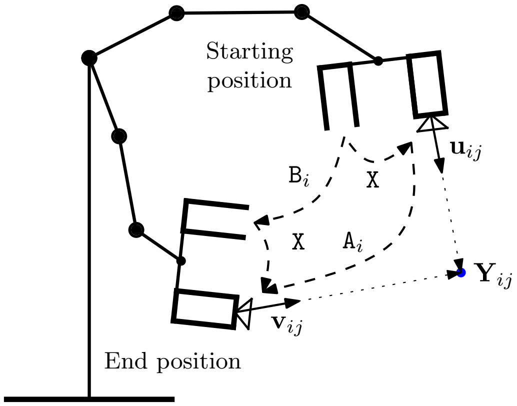


Figure 1: A relative movement of the camera–gripper rig.

1. Introduction

In order to relate the measurements made by a camera mounted on a robotic gripper to the gripper's coordinate frame, a homogeneous transformation from the gripper to the camera needs to be determined. This problem is usually called *hand-eye calibration* and has been studied extensively in the past. The earliest solution strategies can be found in [19, 20, 16, 2, 13]. These methods separate the translational and the rotational parts of the hand-eye calibration and solve for them separately. An early comparison of these methods was given in [21]. Later on, methods solving for the rotation and the translation simultaneously appeared [23, 7, 3, 22].

The common aspect of all the existing methods is that they do not work with the camera measurements directly, but rather with the camera poses derived from them by other methods. The camera poses are usually estimated by observing a known calibration target. Since the calibration target has known dimensions, camera poses with correct scale can be obtained. However, there are many situations when using an accurately manufactured calibration target is not convenient or is not possible at all. Indeed, using a calibration target in applications such as mobile robotics or endoscopy may be unacceptable due to the restrictions in limited on-board weight or respectively to the strict sanitary conditions.

Andreff *et al* [1] proposed a method for "calibration-less" hand-eye calibration based on the Structure-from-Motion (SfM) approach. The authors employed SfM to recover the unknown camera poses. Since SfM can recover camera poses up to scale, the work introduced an explicit scaling factor to the hand-eye calibration equation. A similar approach was presented in [14], where a scaling factor was included into methods [7] and [3].

Recently, a number of methods providing globally optimal solutions to various geometrical problems of Computer Vision appeared. Specifically, two similar frameworks based on minimization of quasi-convex functions using *Second-Order Cone Programming* (SOCP) were presented in [9, 8]. In [5], Hartley and Kahl extended this approach to include both rotation and translation and introduced a branch-and-bound algorithm for finding relative camera poses optimally in the L_∞-norm using Linear Programming. This result allowed Seo *et al* [15] to revisit the hand-eye calibration problem. However, the work [15] assumed all the translations to be zero and provided globally optimal estimates for such translation-less systems.

In this paper we present a modification to the SfM approach to hand-eye calibration. First, we estimate the rotational part of the hand-eye calibration separately using any convenient method. Next, we use SOCP to estimate the translational part from the original image correspondences and robot measurements. This formulation does not require the scaling factor to be estimated explicitly, but it can be recovered easily if needed. Furthermore, the estimated translation is globally optimal with respect to the reprojection error and the L_∞-norm.

2. Problem Formulation

Suppose a camera has been rigidly attached to a robot's gripper. The objective of hand-eye calibration is to derive a homogeneous transformation

$$\mathtt{X} = \begin{pmatrix} \mathtt{R}_\mathtt{X} & \mathbf{t}_\mathtt{X} \\ \mathbf{0}^\top & 1 \end{pmatrix}, \tag{1}$$

where rotation $\mathtt{R}_\mathtt{X}$ and $\mathbf{t}_\mathtt{X} \in \mathbb{R}^3$ are relating the coordinate frames of the gripper and the camera, see Figure 1. This can be done by manipulating the gripper into two or more general positions and observing a scene from different camera viewpoints. In the rest of this paper we will assume that the internal calibration of the camera is known and that the camera measurements are unit vectors representing the directions from the centers of the cameras to the respective 3D points. It has been observed [20] that two motions with non-parallel rotation axes are a sufficient condition for a unique solution for $\mathtt{X}$.

Now, let's suppose that the gripper has been manipulated into n relative movements with the camera measuring m correspondences $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$, $j = 1, \ldots, m$ for every movement $i = 1, \ldots, n$. The restriction for the equal number of correspondences m for every movement is used here just to simplify the notation and can be easily removed by adding another level of indexing. Let $\mathtt{B}_i$ denote the transformation from the coordinate frame of the gripper in its starting position to the coordinate system of the gripper's end position. Transformations $\mathtt{B}_i$ can be obtained from the robot's positioning software and are thus considered to be known. Let $\mathtt{A}_i$ denote the relative camera pose transformations. Camera poses with correct scale can be determined by camera calibration with a known calibration target, or up to scaling using SfM approach. Transformations $\mathtt{A}_i$,$\mathtt{B}_i$ and $\mathtt{X}$ are connected by the following relation, see Figure 1:

$$\mathtt{A}_i\mathtt{X} = \mathtt{X}\mathtt{B}_i. \tag{2}$$

This equation can be easily decomposed into rotational and translational parts

$$\begin{aligned} \mathtt{R}_{\mathtt{A}_i}\mathtt{R}_\mathtt{X} &= \mathtt{R}_\mathtt{X}\mathtt{R}_{\mathtt{B}_i}, && (3)\\ \mathtt{R}_{\mathtt{A}_i}\mathbf{t}_\mathtt{X} + s\mathbf{t}_{\mathtt{A}_i} &= \mathtt{R}_\mathtt{X}\mathbf{t}_{\mathtt{B}_i} + \mathbf{t}_\mathtt{X}, && (4) \end{aligned}$$

where $\mathtt{R}_{\mathtt{A}_i}, \mathtt{R}_{\mathtt{B}_i} \in SO(3)$, $\mathbf{t}_{\mathtt{A}_i}, \mathbf{t}_{\mathtt{B}_i} \in \mathbb{R}^3$ capture the respective rotations and translations. If both $\mathbf{t}_{\mathtt{A}_i}$ and $\mathbf{t}_{\mathtt{B}_i}$ were measured using the same unit, Equation 4 would hold for scaling factor $s = 1$. Note that Equation 3 can be solved for $\mathtt{R}_\mathtt{X}$ regardless of the value of s.

3. Structure from Motion

Structure-from-Motion is a general method for obtaining camera poses from images and consists of the following steps: (i) salient image feature detection and description, (ii) feature descriptor matching between image pairs, (iii) robust pairwise epipolar geometry estimation, and (iv) 3D point triangulation and transformation of the relative camera poses to a common coordinate frame. In this paper, we are mostly interested in steps (ii) and (iii) as the verified matches $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$ and the relative camera rotations $\mathtt{R}_{\mathtt{A}_i}$ are the input to the presented hand-eye calibration method. Having internally calibrated cameras, the decomposition of the obtained essential matrix $\mathtt{E}_{\mathtt{A}_i}$ into $\mathtt{R}_{\mathtt{A}_i}$ and $\mathbf{t}_{\mathtt{A}_i}$ is simple, as the position of the triangulated 3D points can be used to select the correct configuration from the four possible choices [6].

4. Second-Order Cone Programming

It was observed in [8] that various problems from multiview geometry can be written in the following min-max form

$$\min_{\mathbf{x}} \max_{i} \frac{\|\mathtt{F}_i\mathbf{x} + \mathbf{b}_i\|_2}{\mathbf{c}_i^\top\mathbf{x} + d_i} \text{ subject to } \mathbf{c}_i^\top\mathbf{x} + d_i \geq 0, \tag{5}$$

where $\mathbf{x}$ is the vector of unknowns to minimize over, i is the number of measurements, $\mathtt{F}_i$ matrices, and $\mathbf{b}_i, \mathbf{c}_i^\top$ vectors, all of compatible dimensions. If we consider the individual functions $\|\mathtt{F}_i\mathbf{x} + \mathbf{b}_i\|_2 \,/\, \left(\mathbf{c}_i^\top\mathbf{x} + d_i\right)$ as the components of a vector, Problem 5 may be thought of as L_∞-norm minimization of this vector. Problem 5 can also be formulated in the following, more convenient, form:

$$\begin{array}{ll} \text{Minimize} & \gamma \\ \text{subject to} & \|\mathtt{F}_i\mathbf{x} + \mathbf{b}_i\|_2 - \gamma\left(\mathbf{c}_i^\top\mathbf{x} + d_i\right) \leq 0 \end{array} \tag{6}$$

Note that because each of the constraints is convex and the objective function is linear, Problem 6 has a unique solution.

The key observation in [8] is that since for a fixed $\gamma \geq 0$ the constraint $\|\mathtt{F}_i\mathbf{x} + \mathbf{b}_i\|_2 - \gamma\left(\mathbf{c}_i^\top\mathbf{x} + d_i\right) \leq 0$ is a *Second-Order Cone* constraint, we can formulate the following

Second-Order Cone Programming (SOCP) feasibility problem

$$\begin{array}{rl} \text{Given} & \gamma \\ \text{does there exist} & \mathbf{x} \\ \text{subject to} & \|\mathtt{F}_i\mathbf{x}+\mathbf{b}_i\|_2 - \gamma\left(\mathbf{c}_i^\top\mathbf{x}+d_i\right) \leq 0 \\ \text{for} & \forall i\,? \end{array} \tag{7}$$

To solve the original Problem 6 we can employ a bisection scheme for $\gamma \geq 0$ and evaluate Problem 7 repeatedly for fixed values of γ.

Further, it was also observed in [8] that angular reprojection error can be minimized using this approach. With known internal camera calibration, image measurements may be taken to represent unit direction vectors in space. Given a correspondence $\mathbf{u} \leftrightarrow \mathbf{v}$, $\mathbf{u},\mathbf{v} \in \mathbb{R}^3$ and assuming the angle $\angle(\mathbf{u},\mathbf{v})$ is positive and smaller than $\pi/2$, the reprojection error can be represented as

$$\tan\left(\angle(\mathbf{u},\mathbf{v})\right) = \frac{\sin\angle(\mathbf{u},\mathbf{v})}{\cos\angle(\mathbf{u},\mathbf{v})} = \frac{\left\|[\mathbf{u}]_\times\,\mathbf{v}\right\|_2}{\mathbf{u}^\top\mathbf{v}}, \tag{8}$$

where matrix notation $[\mathbf{u}]_\times\,\mathbf{v}$ represents the cross-product $\mathbf{u}\times\mathbf{v}$ [6].

5. Theory

In this section, we derive a modification of Problem 7 and use it to formulate a bisection scheme to estimate the translational part $\mathbf{t}_{\mathtt{X}}$ of hand-eye calibration given a known rotation $\mathtt{R}_{\mathtt{X}}$. Finally, we describe the complete SfM algorithm for hand-eye calibration.

5.1. Feasibility Test

As we can see from Equation 2, the relative camera pose for the i-th relative rig movement can be expressed in robot measurements, rotation $\mathtt{R}_{\mathtt{X}}$, and translation $\mathbf{t}_{\mathtt{X}}$ as

$$\begin{array}{rcl} \mathtt{A}_i & = & \mathtt{X}\mathtt{B}_i\mathtt{X}^{-1} = \begin{pmatrix} \mathtt{R}_{\mathtt{A}_i} & s\mathbf{t}_{\mathtt{A}_i} \\ \mathbf{0}^\top & 1 \end{pmatrix}, \\ \mathtt{R}_{\mathtt{A}_i} & = & \mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top, \\ s\mathbf{t}_{\mathtt{A}_i} & = & \left(\mathtt{I}-\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\right)\mathbf{t}_{\mathtt{X}} + \mathtt{R}_{\mathtt{X}}\mathbf{t}_{\mathtt{B}_i}. \end{array} \tag{9}$$

In was observed in [8, 5] that knowing a relative camera rotation $\mathtt{R}$, correspondences $\mathbf{u}_j \leftrightarrow \mathbf{v}_j, j = 1,\ldots,m$ and error bound γ, the bisection scheme can be used to solve for relative camera translation $\mathbf{t}$ using the following feasibility problem formulation

$$\begin{array}{rl} \text{Given} & \mathtt{R},\gamma \\ \text{do there exist} & \mathbf{t},\mathbf{Y}_j \\ \text{subject to} & \angle\left(\mathbf{u}_j,\mathbf{Y}_j\right) \leq \gamma \\ & \angle\left(\mathbf{v}_j,\mathtt{R}\mathbf{Y}_j+\mathbf{t}\right) \leq \gamma \\ \text{for} & j = 1,\ldots,m\,? \end{array} \tag{10}$$

Note that since the pose transformation cannot be applied directly onto the correspondences, scene points $\mathbf{Y}_j \in \mathbb{R}^3$ also need to be recovered.

In order to apply the bisection framework [8] to hand-eye calibration, we use Problem 10 to formulate a related feasibility problem with $\mathbf{t}_{\mathtt{X}}$ as the unknown. By substituing $\mathtt{R} = \mathtt{R}_{\mathtt{A}_i}$ and $\mathbf{t} = s\mathbf{t}_{\mathtt{A}_i}$ from Equation 9 and repeating for all relative rig movements $i = 1,\ldots,n$ we get

$$\begin{array}{rl} \text{Given} & \mathtt{R}_{\mathtt{X}},\gamma \\ \text{do there exist} & \mathbf{t}_{\mathtt{X}},\mathbf{Y}_{ij} \\ \text{subject to} & \angle\left(\mathbf{u}_{ij},\mathbf{Y}_{ij}\right) \leq \gamma \\ & \angle\left(\mathbf{v}_{ij},\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\mathbf{Y}_{ij}+\right. \\ & \quad\left.\left(\mathtt{I}-\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\right)\mathbf{t}_{\mathtt{X}}+\mathtt{R}_{\mathtt{X}}\mathbf{t}_{\mathtt{B}_i}\right) \leq \gamma \\ \text{for} & i = 1,\ldots,n,\ j = 1,\ldots,m\,? \end{array} \tag{11}$$

Again, as a "by-product" of the problem formulation, scene points $\mathbf{Y}_{ij}$ are recovered. Using the angular error formulation from Equation 8 we can formulate equivalent constraints so that they are linear in the optimized variables $\mathbf{t}_{\mathtt{X}}$ and $\mathbf{Y}_{ij}$, making Problem 11 an SOCP feasibility problem solvable by an SOCP solver. Indeed, we can write the constraints of the first type as

$$\begin{array}{rl} & \angle(\mathbf{u}_{ij},\mathbf{Y}_{ij}) \leq \gamma \\ \Leftrightarrow & \dfrac{\|\mathbf{u}_{ij}\times\mathbf{Y}_{ij}\|_2}{\mathbf{u}_{ij}^\top\mathbf{Y}_{ij}} \leq \tan(\gamma) \\ \Leftrightarrow & \left\|[\mathbf{u}_{ij}]_\times\,\mathbf{Y}_{ij}\right\|_2 - \tan(\gamma)\,\mathbf{u}_{ij}^\top\mathbf{Y}_{ij} \leq 0. \end{array} \tag{12}$$

Analogously, for the second type of constraints we get

$$\begin{array}{rl} & \angle\left(\mathbf{v}_{ij},\mathtt{R}_{\mathtt{A}_i}\mathbf{Y}_{ij}+s\mathbf{t}_{\mathtt{A}_i}\right) \leq \gamma \\ \Leftrightarrow & \dfrac{\|\mathbf{v}_{ij}\times(\mathtt{R}_{\mathtt{A}_i}\mathbf{Y}_{ij}+s\mathbf{t}_{\mathtt{A}_i})\|_2}{\mathbf{v}_{ij}^\top\left(\mathtt{R}_{\mathtt{A}_i}\mathbf{Y}_{ij}+s\mathbf{t}_{\mathtt{A}_i}\right)} \leq \tan(\gamma) \\ \Leftrightarrow & \Big\|\left([\mathbf{v}_{ij}]_\times\,\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\mathbf{Y}_{ij}+[\mathbf{v}_{ij}]_\times\left(\mathtt{I}-\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\right)\mathbf{t}_{\mathtt{X}}\right) \\ & +[\mathbf{v}_{ij}]_\times\,\mathtt{R}_{\mathtt{X}}\mathbf{t}_{\mathtt{B}_i}\Big\|_2 - \\ & \tan(\gamma)\left(\left(\mathbf{v}_{ij}^\top\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\mathbf{Y}_{ij}+\mathbf{v}_{ij}^\top\left(\mathtt{I}-\mathtt{R}_{\mathtt{X}}\mathtt{R}_{\mathtt{B}_i}\mathtt{R}_{\mathtt{X}}^\top\right)\mathbf{t}_{\mathtt{X}}\right)\right. \\ & \left.+\mathbf{v}_{ij}^\top\mathtt{R}_{\mathtt{X}}\mathbf{t}_{\mathtt{B}_i}\right) \leq 0. \end{array} \tag{13}$$

As can be observed from the formulation of Problem 11, since the actual values of translations $\mathbf{t}_{\mathtt{A}_i}$ are not used, the scaling factor s does not need to be known. The correct scale of $\mathbf{t}_{\mathtt{X}}$ is derived solely from $\mathbf{t}_{\mathtt{B}_i}$. However, the value of s can be computed using Equation 9 if needed.

5.2. Bisection

In order to use Problem 11 for the estimation of the translation $\mathbf{t}_{\mathtt{X}}$, a bisection scheme is employed. In every iteration, the value of γ is fixed and Problem 11 is solved. The algorithm starts with $\gamma = \tan(\pi/4)$ and the iteration loop ends once the difference of the lower and upper bounds $\gamma_{\text{low}}, \gamma_{\text{high}}$ reaches a prescribed accuracy ϵ.

Algorithm 1 Bisection

```
Require: $\mathtt{R_X}, \epsilon > 0$
  $\gamma_{\text{low}} \leftarrow 0$
  $\gamma_{\text{high}} \leftarrow 2$
  while $(\gamma_{\text{high}} - \gamma_{\text{low}}) \geq \epsilon$ do
    $\gamma \leftarrow (\gamma_{\text{high}} + \gamma_{\text{low}})/2$
    $(\mathbf{t}_{\mathtt{X}}, \textit{feasible}) \leftarrow$ Feasibility Problem 11
    if feasible then
      $\gamma_{\text{high}} \leftarrow \gamma$
    else
      $\gamma_{\text{low}} \leftarrow \gamma$
    end if
  end while
  return $\mathbf{t}_{\mathtt{X}}$
```

5.3. SfM Algorithm for Hand-Eye Calibration

Finally, we can formulate the complete algorithm for the hand-eye calibration using our SfM approach. Since both the SfM method and the method for $\mathtt{R_X}$ estimation can be changed at the user's convenience, this formulation can be seen as a *meta*-algorithm.

Algorithm 2 SfM Hand-Eye Calibration

1. Estimate the relative camera rotations $\mathtt{R}_{\mathtt{A}_i}$ using a convenient SfM method, *e.g.*, method [17].
2. Estimate $\mathtt{R_X}$ using $\mathtt{R}_{\mathtt{A}_i}$ and $\mathtt{R}_{\mathtt{B}_i}$, *e.g.*, using method [13].
3. Use Algorithm 1 to find the optimal $\mathbf{t}_{\mathtt{X}}$ using $\mathtt{R_X}$ and required precision ϵ.

6. Experimental Results

In the following sections, the proposed Algorithm 2 is validated both by synthetic and real data experiments. We used method [13] to obtain $\mathtt{R_X}$ from $\mathtt{R}_{\mathtt{A}_i}$ and $\mathtt{R}_{\mathtt{B}_i}$ and SeDuMi [18] as the SOCP solver. All the reported times were achieved using a standard Intel Core 2 based consumer PC running 64-bit Linux and MATLAB 7.6.

6.1. Synthetic-data Experiment

A synthetic scene consisting of 100 3D points randomly generated into a ball having radius $1,000\,\mathrm{mm}$ and 10 absolute camera poses set such that the cameras faced approximately the center of the ball were created. The generated 3D points were measured in the respective cameras giving raise to correspondences $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$. Two experiments were conducted with the generated scene. First, the accuracy and efficiency of Algorithm 1 was studied for different values of the prescribed accuracy ϵ. Secondly, the performance of Algorithm 2 was tested on noised correspondences.

ACCURACY experiment Twenty random transformations $\mathtt{X}$ were generated and the tasks composed of the known $\mathtt{R_X}$, the known correspondences $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$, and 9 relative movements $\mathtt{B}_i$, computed from the known absolute camera poses and the generated transformations, were constructed for each of them. These tasks were solved for 5 different values of ϵ ranging from 10^{-6} to 10^{-2} with equal steps on a logarithmic scale and the output was plotted to Figure 2.

The results show that not only the optimized maximum angular reprojection error but also the error of the estimated $\mathbf{t}_{\mathtt{X}}$, measured as the distance between the known and the estimated value of $\mathbf{t}_{\mathtt{X}}$, decreases rapidly with decreasing ϵ. On the other hand, the convergence time increases because more iterations are needed for the bisection.

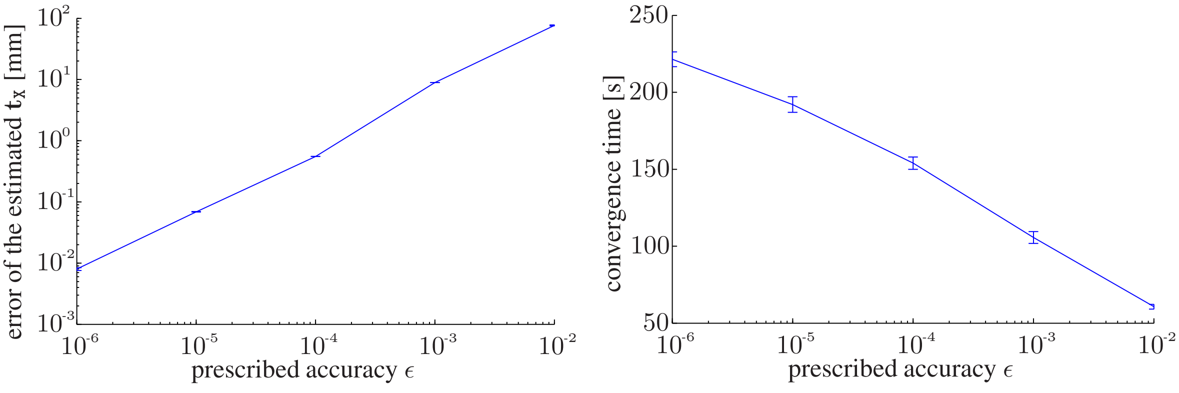
NOISE experiment The same twenty random transformations $\mathtt{X}$ and the corresponding relative movements $\mathtt{B}_i$ were used in the second experiment, where ϵ was fixed to 10^{-5} but the measurements $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$ were corrupted with Gaussian noise in the angular domain. 11 noise levels were used, $\sigma^2 \in \langle 0, 10^{-3} \rangle$ in 10^{-4} steps.

The experiment consisted of two parts, the first one testing the stability of the computation of $\mathbf{t}_{\mathtt{X}}$ for known $\mathtt{R_X}$ by assigning relative camera rotations $\mathtt{R}_{\mathtt{A}_i}$ to the known values while the latter one testing the stability of the whole proposed algorithm. Relative camera rotations were computed by decomposing the essential matrices $\mathtt{E}_{\mathtt{A}_i}$, describing relative camera poses up to scaling, robustly computed from the noised correspondences by RANSAC [4] using the 5-point minimal relative pose problems for calibrated cameras [12] in the latter part.

Figure 3a shows the relation of the maximum angular reprojection error achieved by Algorithm 2 for the various values of σ^2 with $\epsilon = 10^{-5}$ both for the known and for the computed $\mathtt{R_X}$. The relation of the error of the estimated $\mathbf{t}_{\mathtt{X}}$ for the various σ^2 values for both parts of the test can be seen in Figure 3b. As the mean length of the generated $\mathbf{t}_{\mathtt{X}}$ was $259.1\,\mathrm{mm}$ in our experiment, the error of the estimation stayed under 5% even for high noise levels. The convergence times of the algorithm were around 3 minutes for the noise-free task and increasing towards 4 minutes for the tasks using noised correspondences.

6.2. Real-data Experiment

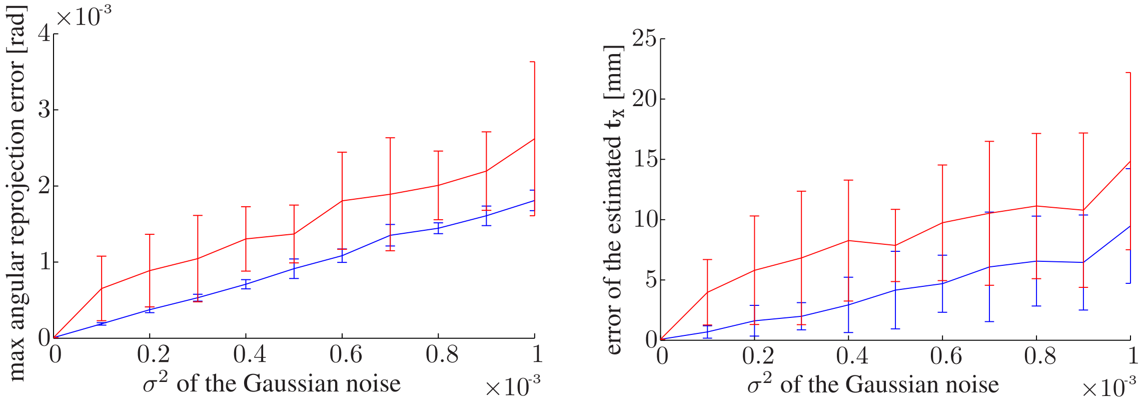
A Mitsubishi MELFA-RV-6S serial manipulator with a Nikon D3000 digital SLR camera and an AF-S DX NIKKOR 18–$55\,\mathrm{mm}$ VR lens (set to $55\,\mathrm{mm}$) was used to acquire the data for the experiment, see Figure 4a. The robot was instructed to move the gripper along the surface of a sphere of radius approximately $700\,\mathrm{mm}$ centered in the middle of the scene objects. The position of the gripper was adjusted using the yaw and pitch angles measured from the center of the sphere to reach ten different locations with five



(a)

(b)

Figure 2: ACCURACY experiment. (a) The mean error of the estimated $\mathbf{t}_{\mathtt{X}}$ for the various values of ϵ plotted in loglog scale together with the variance over the twenty constructed tasks. (b) The mean convergence time of Algorithm 1 for the various values of ϵ plotted in semilogx scale together with the indicated variance.



(a)

(b)

Figure 3: NOISE experiment. (a) The maximum angular reprojection error for the various values of σ^2 recovered by Algorithm 2 with $\epsilon = 10^{-5}$. Blue: Mean maximum error for the known $\mathtt{R}_{\mathtt{X}}$ together with the variance over the twenty tasks. Red: Mean maximum error for $\mathtt{R}_{\mathtt{X}}$ computed by [13] from $\mathtt{R}_{\mathtt{A}_i}$ from the noised correspondences $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$ together with the indicated variance. (b) The error of the estimated $\mathbf{t}_{\mathtt{X}}$ for the various values of σ^2, see (a) for the description of the colors.

different yaw angles for each of the two possible pitch angles. The gripper was set to face the center of the sphere up to a small additive noise. The camera was set to manual mode and images of $3,872 \times 2,592$ pixels were taken using a remote trigger.

Two image sets for two different scenes were acquired—a scene with a calibration target used for obtaining internal camera calibration and a scene with general objects to show the contribution of the proposed method over the hand-eye calibration approaches that rely on a known calibration target. The calibration matrix together with two parameters of radial distortion were computed using [11] and images were radially undistorted prior being further used in order to improve SfM results. Knowing the focal length of the camera, the angular resolution of the acquired images could be computed as $1\,\mathrm{pixel} \approx 1.15 \times 10^{-4}\,\mathrm{rad}$.

CALIBRATION scene Scene CALIBRATION, see Figure 5a-c, was primarily used for internal camera calibration but since the calibration procedure outputs the $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$ correspondences as its by-product, we used the scene for a hand-eye calibration experiment as well. The approach used for the NOISE experiment with the synthetic data was also used to obtain relative camera rotations $\mathtt{R}_{\mathtt{A}_i}$ from the correspondences. Relative robot rotations $\mathtt{R}_{\mathtt{B}_i}$ and translations $\mathbf{t}_{\mathtt{B}_i}$ were obtained from the known gripper-to-base transformations $\mathtt{T}_{\mathtt{B}_i}$.

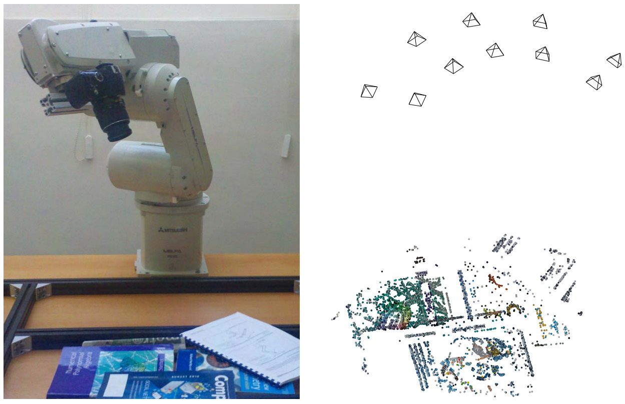
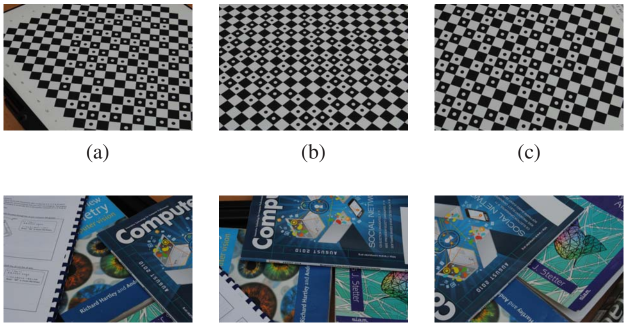

(a) (b)

Figure 4: Real-data experiment. (a) A Mitsubishi MELFA-RV-6S serial manipulator used to acquire the data for the experiment. A Nikon D3000 digital SLR camera mounted on the gripper using a self-made mechanical reduction. (b) The 3D model output from Bundler containing $10,680$ triangulated 3D points and the poses of all the ten cameras.


(a) (b) (c)

(d) (e) (f)

Figure 5: Sample images of our scenes taken by the camera mounted on the gripper of the robot. (a-c) Scene CALIBRATION. (d-f) Scene GENERAL.

A task composed of 9 motions, which were the relative motions between gripper positions 1–2, 2–3, . . . , 9–10, and the respective correspondences $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$, totaling $1,583$ entries was constructed. Algorithm 1 converged in 596 seconds for $\epsilon = 10^{-5}$ giving a solution with the maximum angular error $7.3 \times 10^{-4}\,\mathrm{rad}$. The computed rotation $\mathtt{R}_\mathtt{X}$ was close to the expected, rotation along the z-axis by $-\pi/2$, and the obtained translation from the gripper to the camera center, $-\mathtt{R}_\mathtt{X}^\top \mathbf{t}_\mathtt{X} = (83.5, -19.3, 130.6)^\top$, was close to the result of the method of Tsai [20], $(83.9, -17.8, 133.8)^\top$, and corresponded with a rough physical measurement on the mechanical reduction, $(90, -20, 130)^\top$, showing the validity of the obtained results.

GENERAL scene Scene GENERAL, see Figure 5d-f, was acquired in order to show the performance of the method in real-world conditions. SIFT [10] image features and Bundler [17]—a state-of-the-art open source SfM implementation—were used to obtain the camera poses. Camera focal length from the internal camera calibration was stored as EXIF information in the individual images and Bundler was instructed to preserve the focal lengths read from EXIF.

The resulting 3D model output from Bundler contained $10,680$ triangulated points and the poses of all the ten cameras, see Figure 4b. By examining the 3D point cloud,28 3D points were manually labeled as erroneous and the projections of these points were excluded from the correspondences. This procedure could have been skipped if an SfM method triangulating 3D points from camera triplets instead of pairs was used, since the error rate of such methods is close to zero. Relative camera rotations $\mathtt{R}_{\mathtt{A}_i}$ were computed from the camera projection matrices $\mathtt{P}_{\mathtt{A}_i}$ output from Bundler and relative robot rotations $\mathtt{R}_{\mathtt{B}_i}$ and translations $\mathbf{t}_{\mathtt{B}_i}$ were again obtained from the known gripper-to-base transformations $\mathtt{T}_{\mathtt{B}_i}$. Due to a high number of correspondences, only every tenth member of the set of correspondences $\mathbf{u}_{ij} \leftrightarrow \mathbf{v}_{ij}$ was used for computation giving $1,929$ entries in total for the task composed of the same 9 motions as in the previous experiment.

Algorithm 1 converged in $1,067$ seconds for $\epsilon = 10^{-5}$ giving a solution with the maximum angular error $6.1 \times 10^{-4}\,\mathrm{rad}$. Again, the computed rotation $\mathtt{R}_\mathtt{X}$ was close to the expected one and the obtained translation from the gripper to the camera center, $(97.4, -14.0, 129.9)^\top$, corresponded with the rough physical measurement. We suspect that the difference in the x and y coordinates was caused by the fact that Bundler does not optimize for the position of the principal point and therefore a simplified camera calibration was used, which led to a slightly biased estimation of $\mathtt{R}_{\mathtt{A}_i}$.

7. Conclusion

Using methods of SfM for hand-eye calibration is a natural approach in applications where a precise calibration target cannot be used. However, due to its inherent scale ambiguity, SfM technique brings an additional degree of freedom to the problem. This paper addressed this drawback by formulating the estimation of the hand-eye displacement as an L_∞-norm optimization problem. This formulation recovers the displacement with the correct scale using image correspondences and robot measurements. In addition, optimality of the resulting displacement with respect to the reprojection error is guaranteed. This allowed for the formulation of a novel SfM based hand-eye calibration method which performance was successfully validated by both synthetic and real data experiments.

Acknowledgement

The authors were supported by the EC under projects FP7-SPA-218814 ProVisG and FP7-ICT-247525 HUMAVIPS, by Grant Agency of the CTU Prague project SGS10/277/OHK3/3T/13, by the Czech Government under the research program MSM6840770038, and by the International internship program of the National Institute of Informatics (Tokyo, Japan), Framework of Memorandum of Understanding between the Czech Technical University and National Institute of Informatics. The authors would also like to thank to Martin Meloun and Ondřej Sychrovský for their help with the real-data experiment.

References

[1] N. Andreff, R. Horaud, and B. Espiau. Robot Hand-Eye Calibration using Structure from Motion. *International Journal of Robotics Research*, 20(3):228–248, 2001.

[2] J. C. K. Chou and M. Kamel. Finding the position and orientation of a sensor on a robot manipulator using quaternions. *International Journal of Robotics Research*, 10(3):240–254, 1991.

[3] K. Daniilidis. Hand-eye calibration using dual quaternions. *International Journal of Robotics Research*, 18:286–298, 1998.

[4] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, June 1981.

[5] R. Hartley and F. Kahl. Global optimization through rotation space search. *International Journal of Computer Vision*, 82(1):64–79, 2009.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2003.

[7] R. Horaud and F. Dornaika. Hand-eye calibration. *The International Journal of Robotics Research*, 14(3):195–210, 1995.

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

[9] Q. Ke and T. Kanade. Quasiconvex optimization for robust geometric reconstruction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(10):1834–1847, Oct. 2007.

[10] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.

[11] P. Mareček. A camera calibration system. Master's thesis, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, 2001.

[12] D. Nistér. An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(6):756–770, June 2004.

[13] F. Park and B. Martin. Robot sensor calibration: solving AX=XB on the euclidean group. *Robotics and Automation, IEEE Transactions on*, 10(5):717 –721, Oct. 1994.

[14] J. Schmidt, F. Vogt, and H. Niemann. Calibration-free hand-eye calibration: A structure-from-motion approach. In *DAGM*, pages 67–74, 2005.

[15] Y. Seo, Y.-J. Choi, and S. W. Lee. A branch-and-bound algorithm for globally optimal calibration of a camera-and-rotation-sensor system. In *ICCV*, pages 1173–1178, Sept. 2009.

[16] Y. Shiu and S. Ahmad. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB. *Robotics and Automation, IEEE Transactions on*, 5(1):16–29, Feb. 1989.

[17] N. Snavely, S. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.

[18] J. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999.

[19] R. Tsai and R. Lenz. Real time versatile robotics hand/eye calibration using 3d machine vision. In *ICRA*, pages 554–561 vol.1, Apr. 1988.

[20] R. Tsai and R. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *Robotics and Automation, IEEE Transactions on*, 5(3):345–358, June 1989.

[21] C. Wang. Extrinsic calibration of a vision sensor mounted on a robot. *Robotics and Automation, IEEE Transactions on*, 8:161–175, 1992.

[22] H. Zhang. Hand/eye calibration for electronic assembly robots. *Robotics and Automation, IEEE Transactions on*, 14(4):612–616, Aug. 1998.

[23] H. Zhuang, Z. Roth, and R. Sudhakar. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form $AX = YB$. *Robotics and Automation, IEEE Transactions on*, 10(4):549–554, Aug. 1994.