Solving the Exterior Orientation Problem in 3-D Vision without Initial Guesses

Chien-Ping Lu and Eric Mjolsness

Research Report YALEU/DCS/RR-1033
May 1994

Solving the Exterior Orientation Problem in 3-D Vision without Initial Guesses

Chien-Ping Lu
Department of Computer Science
Yale University
New Haven, CT 06520-8285
Email:lu-chien-ping@cs.yale.edu

Eric Mjolsness
Department of Computer Science
Yale University
New Haven, CT 06520-8285
Email: mjolsness-eric@cs.yale.edu

Abstract

We present a new algorithm for solving the exterior orientation problem. Unlike most of existing methods, it minimizes the 3-D reconstruction error rather than the 2-D projection error. The objective function can be optimized in full by straightforward and efficient coordinate-wise optimizations. An initial guess for camera orientation and position is not required. This algorithm has been tested on synthetic data with varying noise, percentages of outliers, and numbers of data points. We also show the result of applying the alogrithm to hand-eye calibration. Both the accuracy and the speed are very encouraging.

1 Introduction

Given a set of 3-D points, and its 2-D camera view, the problem of determining the rotation and translation that relate the object reference frame to that of the camera is referred to as the *exterior orientation problem* [3] or the *hand-eye calibration problem* [7] when the goal is to locate the camera in object reference frame, and as the *object pose estimation problem* [6] or the *object localization problem* [4] when the goal is to locate the object in the camera reference frame. The rotation and translation are referred to as the *exterior orientation* of the camera or the *pose* of the observed object. It is also an important problem in computer graphics for placement and control of the virtual camera for viewing and rendering [1].

There is extensive work on object pose estimation for 3-D object recognition [8, 4] and photogrammetry [3]. Most existing methods are essentially based on minimizing the error of collinearity equation resulting from fitting the predicted 2-D projections given hypothesized exterior orientation to observed ones. The classic Newton method to minimize this objective function works by iteratively linearizing the collinearity equation around the current approximate solution and solving the linearized system for the next approximate solution. This method usually requires a good starting point. It is reported [6] that for the Newton method to work, the initial approximate solutions have to be within 10% of scale for the translation and within 15^o for each of the three rotation angles.

Instead of doing the nonlinear perspective projection, we can use its inverse, the backprojection, which is linear. It is equivalent to fitting the 3-D points in the camera reference frame to the bundle of the lines of sight associated with the image points. This approach requires estimating the missing depth for each 2-D projection explicitly. It is inappropriate if a classical iterative nonlinear optimization is employed to solved the problem. One algorithm that minimizes 3-D reconstruction error was presented in [2]. It was shown to be globally convergent. No initial approximate solution to true pose was required. However, reasonable initial depths needed to be chosen. It also suffers from slow convergence.

In this paper, we introduce a new depth reconstruction method with which the unknown depth parameters are constrained to a compact set. The new algorithm convergences dramatically faster and is much more accurate. Furthermore, the convergence rate and the accuracy are not affected by the choice of initial depths, as long as they are sufficiently larger than the focal length. The objective function can be augmented with the camera intrinsic parameters to solve the complete camera calibration problem.

2 The Problem Formulation

Given a set of 3-D object coordinates ${}^{\mathbf{o}}\mathbf{P}_i = (X_i, Y_i, Z_i)^{\mathbf{t}}, i = 1, \dots N$, in the object reference frame, and the corresponding coordinates in the camera reference frame ${}^{\mathbf{c}}\mathbf{P}_i$, these two frames can be related by a rigid transformation as

$$(1) \qquad {}^{\mathbf{c}}\mathbf{P}_i = {}^{\mathbf{c}}\boldsymbol{R}_{\mathbf{o}}{}^{\mathbf{o}}\mathbf{P}_i + {}^{\mathbf{c}}\mathbf{T}_{\mathbf{o}},$$

where

$$(2) \qquad {}^{\mathbf{c}}\boldsymbol{R}_{\mathbf{o}} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{pmatrix}, \qquad {}^{\mathbf{c}}\mathbf{T}_{\mathbf{o}} = \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

are a rotation matrix and a translation vector, respectively.

Define $\mathbf{p}_i = (x_i, y_i, f)^{\mathbf{t}}$, $i = 1, \dots N$, to be the perspective projection of ${}^{\mathbf{c}}\mathbf{P}_i$ on the image plane, where f is the focal length. Using the pinhole imaging model, $\mathbf{p}_i$ satisfies the collinearity equation

$$(3) \qquad x_i = f\frac{\mathbf{r}_1 X_i + T_x}{\mathbf{r}_3 Z_i + T_z}, \quad y_i = f\frac{\mathbf{r}_2 Y_i + T_y}{\mathbf{r}_3 Z_i + T_z},$$

or

$$\mathbf{p}_i = \mathcal{P}({}^{o}\mathbf{P}_i; {}^{c}\boldsymbol{R}_{o}, {}^{c}\mathbf{T}_{o}). \tag{4}$$

Equation (4) serves as a basis for most algorithms. The problem can be formulated as that of minimizing the error of collinearity equation

$$\sum_i w_i \|\mathbf{p}_i - \mathcal{P}({}^{o}\mathbf{P}_i; {}^{c}\boldsymbol{R}_{o}, {}^{c}\mathbf{T}_{o})\|^2 \tag{5}$$

with respect to ${}^{c}\boldsymbol{R}_{o}$, ${}^{c}\mathbf{T}_{o}$.

The depth reconstruction is equivalent to reconstructing the 3-D coordinate in the camera reference frame, ${}^{c}\mathbf{P}_i$, which is constrained to lie on the line-of-sight defined by the ideal image point $\mathbf{p}_i$ by

$${}^{c}\mathbf{P}_i = d_i \mathbf{p}_i \quad \text{for } d_i \in \mathbb{R}^+, \tag{6}$$

where fd_i is the depth of the object point in the camera reference frame. The lines of sight may not pass through the object points due to distortion and noise. We seek to minimize the error of such 3-D reconstruction

$$E = \sum_i w_i \|{}^{c}\boldsymbol{R}_{o}{}^{o}\mathbf{P}_i + {}^{c}\mathbf{T}_{o} - d_i \mathbf{p}_i\|^2 \tag{7}$$

with respect to ${}^{c}\boldsymbol{R}_{o}, {}^{c}\mathbf{T}_{o}$, and $\{d_i\}$.

Either of the exterior orientation and the extra depth variables can be solved in closed form given parameters in the other group. The overall solution can be cheaply obtained by coordinate-wise optimization over each group of parameters iteratively. General nonlinear optimization methods are not needed. The coordinate-wise optimization algorithm can be concisely expressed by formulating the objective function as a clocked objective function [9], which is optimized over distinct groups of variables in phases [1]

$$E_{\text{clocked}} = E\langle ({}^{c}\boldsymbol{R}_{o}, {}^{c}\mathbf{T}_{o})^A, \{d_i\}^A \rangle_{\oplus}. \tag{10}$$

Within coordinate-wise optimization on each group of parameters, each parameter, say x, in other groups is clamped, as denoted by $\bar{x}$.

2.1 Solving for the exterior orientation

Find ${}^{c}\boldsymbol{R}_{o}$ and ${}^{c}\mathbf{T}_{o}$ that minimizes

$$E = \sum_i w_i \|{}^{c}\boldsymbol{R}_{o}{}^{o}\mathbf{P}_i + {}^{c}\mathbf{T}_{o} - \bar{d}_i \mathbf{p}_i\|^2, \tag{11}$$

[1] Notations (to be employed recursively):

$$E\langle x, y, \ldots \rangle_{\oplus} : \text{ coordinate-wise optimization of } E \text{ on } x, \text{ then } y, \ldots, \text{ iteratively.} \tag{8}$$

$$x^A : \text{ solving for } x \text{ analytically.} \tag{9}$$

which is a 3-D-3-D pose estimation problem or a absolute orientation problem. When the exterior orientation is represented by an affine transformation with orthonormality constraint on the 3-by-3 matrix, the problem can be solved in closed form by using singular value decomposition [2]. The exterior orientation can also be represented by a dual number quaternion which corresponds to a screw coordinate transform, in which case, the problem can be solved in closed form by computing the eigenvectors for a particular 4-by-4 matrix [10].

In our implementation, the 3-D-3-D pose estimation problem is solved using the method of [10].

2.2 Solving for the depth parameters

Find $\{{}^{c}\mathbf{P}_i\}$ that minimize

$$E = \sum_i w_i \|{}^{c}\bar{\boldsymbol{R}}_{o}{}^{o}\mathbf{P}_i + {}^{c}\bar{\mathbf{T}}_{o} - {}^{c}\mathbf{P}_i\|^2 \tag{12}$$

subject to

$$ {}^{c}\mathbf{P}_i \in \{d_i \mathbf{p}_i | d_i \in \mathbb{R}\} \tag{13}$$

$$\sum_i \|{}^{c}\mathbf{P}_i - \widehat{{}^{c}\mathbf{P}}\|^2 = \sum_i \|{}^{o}\mathbf{P}_i - \widehat{{}^{o}\mathbf{P}}\|^2 \tag{14}$$

with respect to $\{{}^{c}\mathbf{P}_i\}_{i=1,\ldots,N}$. (13) constraints each coordinate in the camera reference frame to lie on the line of sight, therefore the remaining degree of freedom is one, i.e., the depth parameter. (14) is actually a weak rigidity constraint that requires the second-order moment of the set of coordinates in the camera reference frame to be equal to that in the object reference frame. $\widehat{{}^{o}\mathbf{P}}$ and $\widehat{{}^{c}\mathbf{P}}$ can be any linear combination of the coordinates in each reference frame, respectively. We choose to use the means.

The feasible set corresponding to (13) is a N dimensional subspace of $\mathbb{R}^{3N}$, which is convex. If we always start the search from distant coordinates, (14) can be replaced by an inequality

$$\sum_i \|{}^{c}\mathbf{P}_i - \widehat{{}^{c}\mathbf{P}}\|^2 \leq \sum_i \|{}^{o}\mathbf{P}_i - \widehat{{}^{o}\mathbf{P}}\|^2 \tag{15}$$

which makes the corresponding feasible set convex. Now the problem becomes a least squares problem on the intersection of two convex feasible sets, which can be solved by projecting the unconstrained optimum ${}^{c}\bar{\boldsymbol{R}}_{o}{}^{o}\mathbf{P}_i + {}^{c}\bar{\mathbf{T}}_{o}$ on the feasible sets corresponding to (13) and (14) in turn as:

$${}^{c}\mathbf{P}_i^{-} = \frac{({}^{c}\boldsymbol{R}_{o}{}^{o}\mathbf{P}_i + {}^{c}\mathbf{T}_{o})^{t}\mathbf{p}_i}{\mathbf{p}_i^{t}\mathbf{p}_i}\mathbf{p}_i, \tag{16}$$

and

$${}^{c}\mathbf{P}_i = \sqrt{\frac{\sum_i \|{}^{o}\mathbf{P}_i - \widehat{{}^{o}\mathbf{P}}\|^2}{\sum_i \|{}^{c}\mathbf{P}_i^{-} - \widehat{{}^{c}\mathbf{P}^{-}}\|^2}}\,{}^{c}\mathbf{P}_i^{-}. \tag{17}$$

3 Experiments

3.1 Synthetic Data

To demonstrate the robustness of the algorithm, we perform extensive experiments on synthetic data with varying number of points, noise, and percentages of outliers.

A set of 3-D points for $\{^{\mathrm{o}}\mathbf{P}_i\}$ are generated uniformly within a box defined by X_i, Y_i, Z_i $\in$ $[-75, 75]$. The three rotation angles for $^{\mathrm{c}}\boldsymbol{R}_{\mathrm{o}}$ are uniformly selected from $[20, 70]$. T_x and T_y are uniformly selected from $[50, 75]$, and T_z from $[250, 300]$ for $^{\mathrm{c}}\mathbf{T}_{\mathrm{o}}$. The set of 3-D coordinates in the camera reference frame $^{\mathrm{c}}\mathbf{P}_i = {^{\mathrm{c}}\boldsymbol{R}_{\mathrm{o}}}{^{\mathrm{o}}\mathbf{P}_i} + {^{\mathrm{c}}\mathbf{T}_{\mathrm{o}}}$ are generated according to the following control parameters:

Number of points N.

Signal-to-noise ratio SNR. A Gaussian noise $\mathcal{N}(0, \sigma)$ is added to both coordinates of the perspective projection of each $^{\mathrm{c}}\mathbf{P}_i$, where the variance σ is related to SNR by $\mathrm{SNR} = -20 \log \sigma$ dB.

Percentage of outliers PO. A fraction (= PO %) of the 3-D points are selected as outliers. Each of such points $^{\mathrm{c}}\mathbf{P}_i = (^{\mathrm{c}}X_i, {^{\mathrm{c}}Y_i}, {^{\mathrm{c}}Z_i})^t$ is replaced with another 3-D point $(X_i', Y_i', Z_i')^t$, where X_i' and Y_i' are uniformly distributed within $[T_x - 5, T_x + 5]$ and $[T_y - 5, T_y + 5]$, respectively, and $Z_i' = {^{\mathrm{c}}Z_i}$.

The preprocessed 3-D points are then perspectively projected onto the image plane. The focal length is set to 10 mm.

In addition to controlling the parameters described above, we also control the initial depth d for each run.

The following four experiments were conducted:

E1 Set $N = 20, d = 10000$. Estimate the mean of three rotation angle error against SNR (20 dB-80 dB in 10 dB step) for different PO (0 % to 20 % in 5 % step).

E2 Set SNR = 40 dB, $d = 10000$. Estimate the mean of three rotation angle errors against PO (0 %-20 % in 5 % step) for different N (10 to 50 by step of 10).

E3 Set PO = 10 %, $d = 10000$. Estimate the mean of three rotation angle errors against SNR (20 dB-80 dB in 10 dB step) for different N (10 to 50 by step of 10).

E4 Set $N = 20, \mathrm{PO} = 0$. Estimate the mean of three rotation angle errors and the number of iterations (including associated error bars) against $\log_{10} d$ (from 1 to 5 by step of 1).

All the experiments were conducted on a Silicon Graphics IRIS Indigo with MIPS R4400 processor. The average CPU time for each run, including the time for generating the synthetic data set, is as follows:

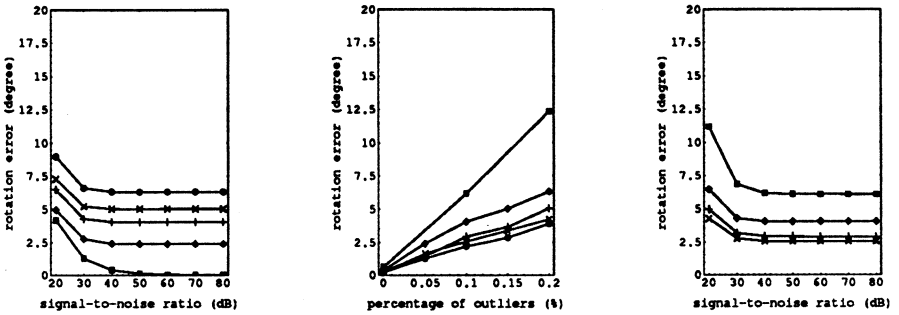

Figure 1: Illustrate performance characteristics of the algorithm with respects to number of points N, signal-to-noise ratio SNR, and percentage of outliers PO. (*left*) N = 20. PO:■ = 0, ♦ = 5%, + = 10%, × = 15%, ∘ = 20%. (*middle*) SNR = 40 dB. N: ■ = 10, ♦ = 20, + = 30, × = 40, ∘ = 50. (*right*) PO = 10%. N: ■ = 10, ♦ = 20, + = 30, × = 40, ∘ = 50. Each data point represents 1,000 trials.

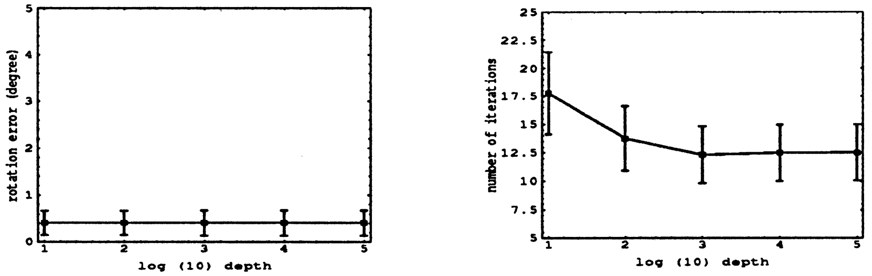

Figure 2: Illustrate insensitivity of the algorithm to initial depth d. Each data point represents 1,000 trials.

number of points	average CPU time (sec)
10	0.056
20	0.075
30	0.089
40	0.10
50	0.12

The first estimate of the exterior orientation is computed from the initial 3-D-3-D pose estimation problem. An initial approximate solution to the pose is not required. Only the initial depth has to be chosen. From the result of **E4**, we found that the performances and the numbers of iterations are not affected by the choices of initial depths, as long as they are several magnitudes larger than the focal length. Actually, with this open-ended requirement on the initial depths, the algorithm can be thought of as initialization-free.

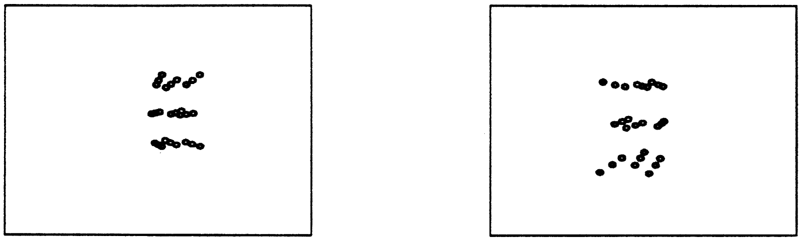

Figure 3: The projections of 27 calibration points as seen through: (*left*) Cohu camera and (*right*) Sony camera.

Notice that the performance of the algorithm is degraded (gracefully, though) with increasing percentage of outliers. We expect that the performance can be improved by using robust methods. Even more aggressively, we are currently investigating an enhancement that solves the problem when the correspondences between the 3-D points and their 2-D projections are totally unknown.

3.2 Hand-Eye Calibration

Our experimental setting for hand-eye calibration consists of a Zebra Zero robot arm, a Cohu camera with an 8 mm lens, a Sony XC-77 camera with a 12.5 mm lens, and two Imaging Technologies digitizers attached to a Sun Sparc II workstation via a Solflower SBus-VME adapter. The robot arm is programmed to reach each of the predefined 27 calibration points by a marker affixed to its end in turns by using inverse kinematics. The corresponding 2-D projections of the calibration points are acquired by tracking the marker through the camera. The tracking system is more fully described in [5].

Given the 3-D coordinates of the calibration points and their corresponding camera view, we compute the rotoation and translation that relate the coordinate system of the robot arm and that of the camera. Since the true rotation and translation are unknown, we measure the 3-D reconsruction and 2-D projection error given the estimated pose. The results for two cameras are summerized as follows:

camera	3-D error (mm)	2-D error (mm)	CPU time (sec)
Cohu	1.97	0.00027	0.15
Sony XC-77	5.52	0.0018	0.15

The 3-D reconstruction errors are measured by comparing the calibration points to the 3-D points reconstructed using the estimated pose and depths. The 2-D projection errors are measured by comparing the observed 2-D projections to those obtained by projecting the calibration points to the image plane.

4 Conclusions

Most of previous methods for solving the exterior orientation problem usually require some "guidelines" to figure out a reasonable initial approximate solution to start the nonlinear search. They are also expected to be slow because of the general nonlinear optimizations involved. In other words, the problem is solved only in a limited sense even with extensive computations.

We presented a clean and efficient algorithm that provides a nearly closed-form solution to the problem. With its robustness and efficiency, the problem can be solved for more arbitrary and dynamical settings.

5 Acknowledgments

We would like to thank Greg Hager for providing hand-eye calibration data, and Robert Haralick for helpful discussions on experiment protocols. We also thank Dimitris Tsioutsias for reading the draft. This work is supported by DARPA/ONR under grant N00014-92-J-4048 and the Yale Center for Theoretical and Applied Neuroscience.

References

[1] J. Blinn, *Where am I? What am I looking at?*, IEEE Computer Graphics and Applications (1988), 76–81.

[2] R. M. Haralick et. al., *Pose estimation from corresponding point data*, IEEE Transaction on Systems, Man, and Cybernetics **19** (1989), no. 6, 1426–1446.

[3] S. K. Ghosh, *Analytical photogrammetry*, Pregamon Press, New York, 1988.

[4] W. E. L. Grimson, *Object recognition by computer*, The MIT Press, Cambridge, Massachusetts, 1990.

[5] G. Hager, S. Puri, and K. Toyama, *A framework for real-time window-based tracking using off-the-shelf-hardware*, Tech. Report YALEU/DCS/RR-988, Yale Computer Science Department, October 1993.

[6] R. M. Haralick and L. G. Shapiro, *Computer and robot vision*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1993.

[7] B. K. P. Horn, *Robot vision*, The MIT Press, Cambridge, Massachusetts, 1986.

[8] D. G. Lowe, *Three-dimensional object recognition from single two-dimensional image*, Artificial Intelligence (1987), no. 31, 355–395.

[9] E. Mjolsness and W. L. Miranker, *Greedy Lagrangians for neural networks: Three levels of optimization in relaxation dynamics*, Tech. Report YALEU/DCS/TR-945, Yale Computer Science Department, January 1993.

[10] M. W. Walker and L. Shao, *Estimating 3-D location parameters using dual number quaternions*, CVGIP: Image Understanding **54** (1991), no. 3, 358–367.