

Comparing Two Sets of Corresponding Six Degree of Freedom Data

Mili Shah

*Department of Mathematics and Statistics at Loyola University of Maryland
4501 North Charles Street, Baltimore, MD, 21210*

Abstract

This paper is concerned with comparing two sets of corresponding six degree of freedom data that consist of both object position and object orientation. Specifically, the best rotation and translation that aligns the position and orientation of one data set to the other is constructed by solving an optimization problem. In addition, a statistical method that identifies outliers in the data sets is proposed.

Keywords: Computer Vision; Singular Value Decomposition; Absolute Orientation Problem; Hand-Eye Calibration; Six Degree of Freedom; Translation; Rotation; Camera Calibration

1. Introduction

With the advent of newer and more technologically advanced computer vision systems, there is greater need for mathematical techniques to calibrate these systems. For example, consider an object going down an assembly line. Generally, the assembly line stops in order for robots to operate on the object. But if the robots could visually track the object, then they could operate on it while the assembly line is in motion and thus increase efficiency. In order for the robots to track the object, each is equipped with cameras that incorporate a computer vision system. The goal of this paper is to evaluate the accuracy of a specific computer vision system by comparing the data it gathers with data that are collected from a precise sensor system considered ground truth. The problem with comparing these two data sets is that they are not necessarily in the same coordinate system. Therefore, a transform from the computer vision system's data stream to the coordinate system of ground truth is necessary. Once this transform is obtained, a metric can be calculated to track how well the underlying computer vision system works. By ranking these metrics, the optimal system can be chosen.

I am interested in evaluating computer vision systems that obtain six degree of freedom (6DoF) data which represent both the position and the orientation of an object. It should be noted that the approach presented in this paper can be extended beyond performance evaluation. In general, this approach will create the rotation and translation that best transforms one set of 6DoF data into another. Therefore, this process can be applied to any computer vision problem

Email address: mishah@loyola.edu (Mili Shah)

URL: <http://math.loyola.edu/~mili> (Mili Shah)

Preprint submitted to Computer Vision and Image Understanding

May 25, 2011

where 6DoF data are collected and/or calibrated such as visual servoing, object recognition, and motion estimation.

The 6DoF data represent translations along three perpendicular axes: left and right (along the x axis), forward and backward (along the y axis), and up and down (along the z axis); along with the rotations about those three perpendicular axes (roll r_x , pitch r_y , and yaw r_z). In order to evaluate a given 6DoF computer vision system, independent sets of data are simultaneously collected from the given system and a ground truth sensor. This allows both sets of data to be synchronized and permits the establishment of a correspondence between the two. Typically, each set of data has its own coordinate system. Therefore, a transformation is needed in order to compare the given computer vision system with ground truth. In order to find such a transformation, a matrix representation for the 6DoF data is used. If the 6DoF representation of an object is represented as (x, y, z, r_x, r_y, r_z) , then it may be arranged as a homogeneous matrix

$$\mathbf{H} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix},$$

where $t = (x, y, z)^T$ represents the position of the given object and

$$R = R_x R_y R_z$$

with

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(r_x) & -\sin(r_x) \\ 0 & \sin(r_x) & \cos(r_x) \end{pmatrix}, R_y = \begin{pmatrix} \cos(r_y) & 0 & \sin(r_y) \\ 0 & 1 & 0 \\ -\sin(r_y) & 0 & \cos(r_y) \end{pmatrix}, R_z = \begin{pmatrix} \cos(r_z) & -\sin(r_z) & 0 \\ \sin(r_z) & \cos(r_z) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

represents the orientation of a given object. Given two sets of such corresponding 6DoF data

$$\mathbf{X} = \left[\begin{pmatrix} R_0 & t_0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} R_1 & t_1 \\ 0 & 1 \end{pmatrix}, \dots, \begin{pmatrix} R_{n-1} & t_{n-1} \\ 0 & 1 \end{pmatrix} \right]$$

$$\mathbf{X}' = \left[\begin{pmatrix} R'_0 & t'_0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} R'_1 & t'_1 \\ 0 & 1 \end{pmatrix}, \dots, \begin{pmatrix} R'_{n-1} & t'_{n-1} \\ 0 & 1 \end{pmatrix} \right],$$

the best rotation Ω and translation τ that fits the data is constructed. In other words, the best homogeneous matrix $\mathbf{H} = \begin{pmatrix} \Omega & \tau \\ 0 & 1 \end{pmatrix}$ that minimizes

$$\min_{\mathbf{H}} \|\mathbf{H}\mathbf{X} - \mathbf{X}'\|^2 \quad (1)$$

is constructed.

The solution \mathbf{H} to the minimization problem of Equation (1) involves a two step process:

1. Find the rotation Ω that minimizes

$$\min_{\Omega} \left\| \Omega \begin{pmatrix} R_0 & \bar{t}_0 & \dots & R_{n-1} & \bar{t}_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \bar{t}'_0 & \dots & R'_{n-1} & \bar{t}'_{n-1} \end{pmatrix} \right\|^2 \quad (2)$$

where

$$\bar{t}_i = t_i - t \quad \text{and} \quad t = \frac{1}{n} \sum_{i=0}^{n-1} t_i$$

$$\bar{t}'_i = t'_i - t' \quad \text{and} \quad t' = \frac{1}{n} \sum_{i=0}^{n-1} t'_i.$$

2. Set the best transformation

$$\tau = t' - \Omega t, \quad (3)$$

where Ω is calculated from Step 1.

The simpler problem of finding a closed-form solution to the best rotation and translation fitting two sets of three-dimensional point correspondences (which represents only position and hence has 3DoF) has been around since the 1980's [1, 2]. Most formulations are reduced to finding a rotation Ω and translation τ that solves

$$\min_{\Omega, \tau} \|(\Omega \mathbf{Y} + \tau) - \mathbf{Y}'\|^2 \quad (4)$$

where \mathbf{Y} and \mathbf{Y}' are $3 \times n$ matrices such that the i -th column of \mathbf{Y}' is given by

$$y'_i = \Omega y_i + \tau + e_i,$$

where y_i and y'_i are the i -th column of \mathbf{Y} and \mathbf{Y}' , respectively and e_i is a noise vector. This minimization is commonly known as the *absolute orientation* problem. One of the issues with this approach is that there are certain cases where there are many – if not infinite – solutions to the minimization problem (4) [1]. An example of this case is when all the points lie on the same line such as when an object goes down a linear assembly line. This degeneracy is not a problem with the 6DoF method since the object's orientation is also included in the minimization. This creates additional constraints to the associated 6DoF linear system and thus a unique solution can be found. In Section 5.1, an example will be presented that illustrates this degeneracy.

Historically, there are four main approaches to finding closed form solutions for the 3DoF representation. The first method by Arun, Huang, and Blostein [1] is based on finding the best orthogonal matrix that fits the two sets of data and declares it the rotation. An equivalent method - by Horn, Hilden, and Negahdaripour [2] - looks for the square-root of a symmetric matrix to represent rotation. A problem with both of these methods is that the matrix that is calculated may not necessarily be a rotation (in fact it may be a reflection). Therefore, the results from the algorithm may have to be discarded. In contrast, the method that is presented here is guaranteed to be a rotation matrix. The last two approaches – one by Horn [3] and the other by Walker, Shao, and Volz [4] – are based on quaternions. Modern extensions of the conventional four methods have been formulated by Umeyama [5] and Kanatani [6]. There are also many iterative methods for solving 3DoF systems as suggested in [2]. However, all 3DoF methods only evaluate the position of the object, neglecting orientation.

In this paper, I advance the formulation of the 3DoF representation by not only considering the position but also the orientation data of an object. Therefore, a complete performance evaluation for a given 6DoF computer vision system can be accomplished. The formulation of our work is similar to Govindu's work [7, 8, 9] on estimating the internal parameters of a camera. However, in Govindu's work a direct search method is used to solve a variation of the optimization problem of Equation (1). In contrast, the 6DoF method presented here formulates a closed form solution to the problem.

This closed formed solution is similar to the closed formed solutions for the hand-eye calibration problem $\mathbf{AX} = \mathbf{ZB}$. Here, \mathbf{X} and \mathbf{Z} are unknown homogeneous matrices and \mathbf{A} and \mathbf{B} are known homogeneous matrices [10, 11, 12, 13]. Closed form solutions for the hand-eye calibration method are formulated by separating the problem into its orientational and positional

components by noting that

$$\mathbf{A}\mathbf{X} = \mathbf{Z}\mathbf{B}$$

$$\begin{pmatrix} R_A & t_A \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_X & t_X \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R_Z & t_Z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_B & t_B \\ 0 & 1 \end{pmatrix}.$$

Thus, the orientational component is represented as

$$R_A R_X = R_Z R_B,$$

while the positional component is represented as

$$R_A t_X + t_A = R_Z t_B + t_Z.$$

In this paper, \mathbf{X} is assumed to be known. Thus, the hand-eye calibration is simplified to $\mathbf{A} = \mathbf{Z}\mathbf{B}$ for unknown $\mathbf{Z} = \begin{pmatrix} \Omega & \tau \\ 0 & 1 \end{pmatrix}$. Then using the closed form solution of the hand-eye calibration method, Ω is computed as the rotation that best fits the orientational component, i.e.

$$\Omega R_B = R_A,$$

while the translation τ is found by calculating the least squares solution to

$$\Omega t_B + \tau = t_A$$

where Ω is the rotation calculated solely from the orientational component. In other words, the hand-eye calibration method ignores the positional data when computing the rotation Ω . In contrast, the method formulated in this paper obtains the rotation Ω from both the orientational and positional data. A comparison between the closed form solutions of the hand-eye calibration method, the 3DoF method, and the 6DoF method formulated in this paper is presented on simulated data in Section 5.2 and on real data in Section 5.3.

In this paper, $\|\cdot\|$ denotes the Frobenius norm, so

$$\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^T)} = \sqrt{\text{tr}(\mathbf{A}^T\mathbf{A})}$$

where T denotes the transpose operator. And, $\text{tr}()$ denotes the matrix trace operation, while $\text{diag}(d_1 \dots d_n)$ represents the diagonal matrix with entries $d_1 \dots d_n$.

2. Simplifying Rotation and Translation

Here, I will outline the methodology that reduces the original system (1) to the two-step process shown in Equation (2) and Equation (3). First, observe that

$$\begin{aligned} \|\mathbf{H}\mathbf{X} - \mathbf{X}'\|^2 &= \left\| \begin{pmatrix} \Omega & \tau \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R_0 & t_0 & \dots & R_{n-1} & t_{n-1} \\ 0 & 1 & \dots & 0 & 1 \end{pmatrix} - \begin{pmatrix} R'_0 & t'_0 & \dots & R'_{n-1} & t'_{n-1} \\ 0 & 1 & \dots & 0 & 1 \end{pmatrix} \right\|^2 \\ &= \left\| \begin{pmatrix} \Omega R_0 - R'_0 & \Omega t_0 + \tau - t'_0 & \dots & \Omega R_{n-1} - R'_{n-1} & \Omega t_{n-1} + \tau - t'_{n-1} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \right\|^2 \\ &= \left\| \Omega \begin{pmatrix} R_0 & \dots & R_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \dots & R'_{n-1} \end{pmatrix} \right\|^2 + \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2 \end{aligned} \quad (5)$$

Now, let the centroids for the two data sets be given by

$$t = \frac{1}{n} \sum_{i=0}^{n-1} t_i \quad \text{and} \quad t' = \frac{1}{n} \sum_{i=0}^{n-1} t'_i$$

and define

$$\mathbf{T} = \tau + \Omega t - t'.$$

Then for $i = 0, \dots, n-1$

$$\bar{t}_i = t_i - t \quad \text{and} \quad \bar{t}'_i = t'_i - t.$$

Therefore,

$$\begin{aligned} \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2 &= \sum_{i=0}^{n-1} \|\Omega(t_i - t) - (t'_i - t') + \tau + \Omega t - t'\|^2 \\ &= \sum_{i=0}^{n-1} \|\Omega \bar{t}_i - \bar{t}'_i + \mathbf{T}\|^2 \\ &= \sum_{i=0}^{n-1} \|\Omega \bar{t}_i - \bar{t}'_i\|^2 + 2\mathbf{T}^T \left(\sum_{i=0}^{n-1} \Omega \bar{t}_i - \bar{t}'_i \right) + n\|\mathbf{T}\|^2. \end{aligned}$$

Since \bar{t}_i and \bar{t}'_i are mean-adjusted,

$$\sum_{i=0}^{n-1} \bar{t}_i = \sum_{i=0}^{n-1} \bar{t}'_i = 0.$$

Thus,

$$\sum_{i=0}^{n-1} \Omega \bar{t}_i - \bar{t}'_i = 0,$$

and

$$\sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2 = \sum_{i=0}^{n-1} \|\Omega \bar{t}_i - \bar{t}'_i\|^2 + n\|\mathbf{T}\|^2. \quad (6)$$

Moreover, if Equation (5) is minimized then

$$\begin{aligned} \min_{\mathbf{H}} \|\mathbf{H}\mathbf{X} - \mathbf{X}'\|^2 &= \min_{\Omega, \tau} \left\| \Omega \begin{pmatrix} R_0 & \dots & R_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \dots & R'_{n-1} \end{pmatrix} \right\|^2 + \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2 \\ &= \min_{\Omega, \tau} \left\| \Omega \begin{pmatrix} R_0 & \dots & R_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \dots & R'_{n-1} \end{pmatrix} \right\|^2 + \sum_{i=0}^{n-1} \|\Omega \bar{t}_i - \bar{t}'_i\|^2 + n\|\mathbf{T}\|^2 \\ &= \min_{\Omega, \tau} \left\| \Omega \begin{pmatrix} R_0 & \bar{t}_0 & \dots & R_{n-1} & \bar{t}_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \bar{t}'_0 & \dots & R'_{n-1} & \bar{t}'_{n-1} \end{pmatrix} \right\|^2 + n\|\mathbf{T}\|^2 \end{aligned}$$

Note for any given rotation Ω , $\mathbf{T} = 0$ by defining

$$\tau = t' - \Omega t \Rightarrow \mathbf{T} = \tau + \Omega t - t' = 0.$$

Thus, in order to solve Equation (1), first calculate Ω that minimizes

$$\min_{\Omega} \left\| \Omega \begin{pmatrix} R_0 & \bar{t}_0 & \dots & R_{n-1} & \bar{t}_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \bar{t}'_0 & \dots & R'_{n-1} & \bar{t}'_{n-1} \end{pmatrix} \right\|^2,$$

then set

$$\tau = t' - \Omega t.$$

2.1. Finding Ω

In the previous section, it was shown that finding the best fitting homogeneous transformation matrix is dependent on finding the best rotation that minimizes Equation (2):

$$\min_{\Omega} \left\| \Omega \begin{pmatrix} R_0 & \bar{t}_0 & \dots & R_{n-1} & \bar{t}_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \bar{t}'_0 & \dots & R'_{n-1} & \bar{t}'_{n-1} \end{pmatrix} \right\|^2.$$

For simplicity, this problem will be reformulated to

$$\min_{\Omega} \left\| \Omega \bar{\mathbf{X}} - \bar{\mathbf{X}}' \right\|^2, \quad (7)$$

where the sets which include the mean-adjusted positional data

$$\begin{aligned} \bar{\mathbf{X}} &= \begin{pmatrix} R_0 & \bar{t}_0 & \dots & R_{n-1} & \bar{t}_{n-1} \end{pmatrix} \\ \bar{\mathbf{X}}' &= \begin{pmatrix} R'_0 & \bar{t}'_0 & \dots & R'_{n-1} & \bar{t}'_{n-1} \end{pmatrix}. \end{aligned}$$

However,

$$\left\| \Omega \bar{\mathbf{X}} - \bar{\mathbf{X}}' \right\|^2 = \left\| \bar{\mathbf{X}} \right\|^2 - 2\text{tr}(\Omega \bar{\mathbf{X}} \bar{\mathbf{X}}'^T) + \left\| \bar{\mathbf{X}}' \right\|^2.$$

Therefore, the Ω that solves the minimization problem of Equation (7) is equivalent to the rotation matrix Ω that solves

$$\max_{\Omega} \text{tr}(\Omega \bar{\mathbf{X}} \bar{\mathbf{X}}'^T) \quad (8)$$

There is a plethora of research on finding the best rotation matrix Ω . Most of these methods are based on finding the best orthogonal matrix that fits the data. In most applications, this method works. However, there can be instances where the best orthogonal matrix that is produced could have determinant -1 , meaning that the best orthogonal matrix is not a rotation but actually a reflection. In this section, a method for calculating the best rotational approximation to a set of data that is guaranteed to have determinant 1 will be described. This work reaches the same conclusion as Umeyama's work [5], though the formulation of the proof presented here is much simpler.

In order to construct the best rotation, the following Lemma will be of importance.

Lemma 2.1. For a given 3×3 matrix \mathbf{M} and rotation Ω

$$\text{tr}(\Omega \mathbf{M}) \leq \text{tr}(\mathbf{D} \mathbf{M}), \quad (9)$$

where

$$\mathbf{D} = \begin{cases} \text{diag}(1, 1, 1) & \text{if } \det(\mathbf{V} \mathbf{U}^T) = 1, \\ \text{diag}(1, 1, -1) & \text{if } \det(\mathbf{V} \mathbf{U}^T) = -1 \end{cases}$$

and the full singular value decomposition (SVD) of

$$\mathbf{M} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T.$$

Proof. First notice that

$$\text{tr}(\Omega\mathbf{M}) = \text{tr}(\Omega\mathbf{U}\Sigma\mathbf{V}^T) = \text{tr}(\mathbf{V}^T\Omega\mathbf{U}\mathbf{D}\mathbf{D}\Sigma),$$

since \mathbf{D}^2 is the identity and $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ for matrices \mathbf{A} and \mathbf{B} of appropriate degree. But $\hat{\Omega} = \mathbf{V}^T\Omega\mathbf{U}\mathbf{D}$ is an orthogonal matrix with determinant 1 and hence a rotation matrix. Therefore,

$$\text{tr}(\Omega\mathbf{M}) = \text{tr}(\hat{\Omega}\mathbf{D}\Sigma) \leq \text{tr}(\mathbf{D}\Sigma)$$

■

Moreover, if a rotation Ω can be constructed such that

$$\text{tr}(\Omega\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = \text{tr}(\mathbf{D}\Sigma),$$

then the minimization problem of Equation (7) is solved.

Theorem 2.2. *The solution to the maximization problem of Equation (8) is*

$$\Omega = \mathbf{V}\mathbf{D}\mathbf{U}^T$$

where the full SVD of the 3×3 matrix

$$\bar{\mathbf{X}}\bar{\mathbf{X}}^T = \mathbf{U}\Sigma\mathbf{V}^T$$

and

$$\mathbf{D} = \begin{cases} \text{diag}(1, 1, 1) & \text{if } \det(\mathbf{V}\mathbf{U}^T) = 1, \\ \text{diag}(1, 1, -1) & \text{if } \det(\mathbf{V}\mathbf{U}^T) = -1 \end{cases}$$

Proof. From Lemma 2.1, the maximization problem of Equation (8) is solved if a rotation matrix Ω can be constructed such that

$$\text{tr}(\Omega\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = \text{tr}(\mathbf{D}\Sigma).$$

Let

$$\Omega = \mathbf{V}\mathbf{D}\mathbf{U}^T.$$

Then

$$\text{tr}(\Omega\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = \text{tr}([\mathbf{V}\mathbf{D}\mathbf{U}^T][\mathbf{U}\Sigma\mathbf{V}^T]) = \text{tr}(\mathbf{D}\Sigma).$$

■

Therefore, the optimal homogeneous matrix $\mathbf{H} = \begin{pmatrix} \Omega & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}$ may be constructed by

1. Setting

$$\Omega = \mathbf{V}\mathbf{D}\mathbf{U}^T,$$

where the SVD of

$$\bar{\mathbf{X}}\bar{\mathbf{X}}^T = \mathbf{U}\Sigma\mathbf{V}^T$$

and

$$\mathbf{D} = \begin{cases} \text{diag}(1, 1, 1) & \text{if } \det(\mathbf{V}\mathbf{U}^T) = 1, \\ \text{diag}(1, 1, -1) & \text{if } \det(\mathbf{V}\mathbf{U}^T) = -1 \end{cases}.$$

2. Setting

$$\tau = t' - \Omega t.$$

3. Error Metrics

For many applications, it is beneficial to understand how well the homogeneous matrix \mathbf{H} fits the orientation of the 6DoF data independently of the position of the 6DoF data. Examples of such applications arise in the hand-eye calibration methods that were presented in Section 1.

To separate the data, consider Equation (5),

$$\begin{aligned} \|\mathbf{H}\mathbf{X} - \mathbf{X}'\|^2 &= \left\| \Omega \begin{pmatrix} R_0 & \dots & R_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \dots & R'_{n-1} \end{pmatrix} \right\|^2 + \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2 \\ &= \sum_{i=0}^{n-1} \|\Omega R_i - R'_i\|^2 + \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2. \end{aligned}$$

This is a separation of the orientational data from the positional data. Moreover, once the Ω and τ of the homogeneous matrix \mathbf{H} are calculated from the procedure outlined in Section 2, a means to find how well Ω and τ fit the data can be constructed. Notice that for the orientation

$$\begin{aligned} \|\Omega R_i - R'_i\|^2 &= \|\Omega R_i\|^2 - 2\text{tr}(\Omega R_i R'_i{}^T) + \|R'_i\|^2 \\ &= 6 - 2\text{tr}(\Omega R_i R'_i{}^T) \\ &= 6 - 2(1 + 2\cos\theta) \\ &\leq 8. \end{aligned}$$

since $\|R\|^2 = 3$ and $\text{tr}(R) = 1 + 2\cos\theta$ for any rotation matrix R with eigenvalues $\{1, \cos\theta \pm i\sin\theta\}$. Therefore, if θ is approximately equal to 0, then $6 - 2(1 + 2\cos\theta) \approx 6 - 2(3) = 0$, whereas if $\theta \approx \pi$ then $6 - 2(1 + 2\cos\theta) \approx 6 - 2(-1) = 8$. Therefore, a metric or *percentage of accuracy* to evaluate the orientation for a given homogeneous matrix \mathbf{H} (hence rotation Ω and translation τ) can be calculated as

$$0 \leq 1 - \frac{1}{8} \|\Omega R_i - R'_i\|^2 \leq 1.$$

A metric for the positions can be calculated in a similar way. In this case, the norm

$$\|\Omega t_i + \tau - t'_i\|^2.$$

That is, the closeness of the vector $\Omega t_i + \tau$ to t'_i for a given rotation Ω and translation τ can be constructed. In order to construct a metric or *percentage of accuracy* for this data, consider the dot product of the normalized vectors, i.e.

$$0 \leq \left| \frac{(\Omega t_i + \tau)^T t'_i}{\|\Omega t_i + \tau\| \|t'_i\|} \right| \leq 1.$$

If the angle between the vectors is 0, then the algorithm has 100% accuracy. A point of concern with this method is that the magnitude of the vectors are not taken into consideration. Thus, this metric may exhibit 100% accuracy while the vectors are not exactly equal. As a result, one may want to compare the magnitude of

$$\|\Omega t_i + \tau - t'_i\|$$

with the magnitude of the positions t_i and t'_i to determine the accuracy of the algorithm. However, this metric does not have an upper-bound so it may be difficult to compare the results from different problem sets as is possible with the first metric presented.

4. Outliers

Outliers generally arise when collecting data. In this section, a statistical method is constructed that detects outliers in the data sets. The method is based on the statistical tool known as the Interquartile range (IQR). The IQR defines the difference between the 25th (Q1) and 75th (Q3) percentiles of the data stream. In other words,

$$\text{IQR} = \text{Q3} - \text{Q1}.$$

Definition 4.1. *A point x in the data stream is an outlier if*

$$x \geq \text{Q3} + 1.5 \times \text{IQR}.$$

Using this definition, which is based on Tukey's work [14], a statistical method to detect outliers in the data stream is constructed.

The method begins by constructing the best homogeneous matrix

$$\mathbf{H} = \begin{pmatrix} \Omega & \tau \\ 0 & 1 \end{pmatrix},$$

to fit the two data sets as outlined in Section 2. Then for each set of points

$$\begin{aligned} \mathbf{X} &= \left[\begin{pmatrix} R_0 & t_0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} R_1 & t_1 \\ 0 & 1 \end{pmatrix}, \dots, \begin{pmatrix} R_{n-1} & t_{n-1} \\ 0 & 1 \end{pmatrix} \right] \\ \mathbf{X}' &= \left[\begin{pmatrix} R'_0 & t'_0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} R'_1 & t'_1 \\ 0 & 1 \end{pmatrix}, \dots, \begin{pmatrix} R'_{n-1} & t'_{n-1} \\ 0 & 1 \end{pmatrix} \right], \end{aligned}$$

calculate the error

$$\mathbf{e}_i = \|\mathbf{H}\mathbf{X}_i - \mathbf{X}'_i\|^2.$$

From this collection of \mathbf{e}_i , outliers \mathbf{e}_j are defined using Definition 4.1. The corresponding $\mathbf{X}_j = \begin{pmatrix} R_j & t_j \\ 0 & 1 \end{pmatrix}$ and $\mathbf{X}'_j = \begin{pmatrix} R'_j & t'_j \\ 0 & 1 \end{pmatrix}$ from the data sets \mathbf{X} and \mathbf{X}' , respectively, are thrown out and a new best fitting homogeneous matrix \mathbf{H} is calculated from the updated \mathbf{X} and \mathbf{X}' . In our experiments, one iteration of this method is sufficient to locate outliers. However, the iteration could continue until the norm between the previous and new homogeneous matrix is under a predetermined tolerance or until no outliers are detected.

5. Experiments

5.1. Linear Motion Degeneracy

In this section, the degeneracy of the 3DoF method that occurs when all points lie on the same line is explored. To illustrate, consider the mean-adjusted linear set of positional data

$$\bar{\mathbf{Y}} = \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and its perfectly corresponding set of mean-adjusted points

$$\bar{\mathbf{Y}}' = \Omega \bar{\mathbf{Y}},$$

where Ω is a random rotation matrix. For simplicity we assume that the translation $\tau = 0$. These sets of data are collinear and thus infinite solutions for the optimal rotation matrix to fit the data exist [1]. This is a result of $\bar{\mathbf{Y}} \bar{\mathbf{Y}}'^T$ being a rank-1 matrix. Specifically, the SVD of

$$\bar{\mathbf{Y}} \bar{\mathbf{Y}}' = \sigma_1 u_1 v_1^T + 0u_2 v_2^T + 0u_3 v_3^T$$

where σ_1 is the leading singular value and u_i and v_i are the left and right singular vector respectively for $i = 1, 2, 3$. Since this matrix is rank-1, the second and third singular values are 0, and thus infinite options for the corresponding left and right singular vectors exist. As a result, the optimal rotation matrix that is formulated from these left and right singular vectors is not unique. In contrast, the 6DoF method requires both the positional and orientational data. Thus, the data are represented as

$$\bar{\mathbf{X}} = \begin{pmatrix} & -2 & & -1 & & 0 & & 1 & & 2 \\ R_1 & 0 & R_2 & 0 & R_3 & 0 & R_4 & 0 & R_5 & 0 \\ & 0 & & 0 & & 0 & & 0 & & 0 \end{pmatrix}$$

and its perfectly corresponding set of data

$$\bar{\mathbf{X}}' = \Omega \bar{\mathbf{X}}.$$

Here R_i represents the orientation for points $i = 1, 2, \dots, 5$. Notice that these sets are not collinear since the columns of each orientation R_i are orthogonal and thus full-rank. Therefore, the 6DoF method formulates a unique rotation for this data [1]. It should be noted that the hand-eye calibration method will also formulate a unique rotation since the data consist of only the orientations which again form a non-collinear set. Thus, a unique rotation matrix can be found.

5.2. Comparison of Methods with Simulated Data

In this section, the hand-eye calibration method, the 3DoF method, and the 6DoF method formulated in this paper are explored. Recall, the hand-eye calibration method obtains the rotation Ω by optimizing the orientational data, i.e. by solving

$$\min_{\Omega} \left\| \Omega \begin{pmatrix} R_0 & \dots & R_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \dots & R'_{n-1} \end{pmatrix} \right\|^2.$$

In contrast, the 3DoF method obtains the rotation Ω by optimizing the positional data, i.e. by solving

$$\min_{\Omega, \tau} \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2.$$

Now consider the formulation of Ω from the 6DoF method which is constructed by minimizing Equation (5)

$$\min_{\mathbf{H}} \|\mathbf{H}\mathbf{X} - \mathbf{X}'\|^2 = \min_{\Omega, \tau} \left\| \Omega \begin{pmatrix} R_0 & \dots & R_{n-1} \end{pmatrix} - \begin{pmatrix} R'_0 & \dots & R'_{n-1} \end{pmatrix} \right\|^2 + \sum_{i=0}^{n-1} \|\Omega t_i + \tau - t'_i\|^2.$$

One can easily see that this formulation is just a combination of the hand-eye calibration method and the 3DoF method. In other words, the rotation Ω for the 6DoF method is formulated by

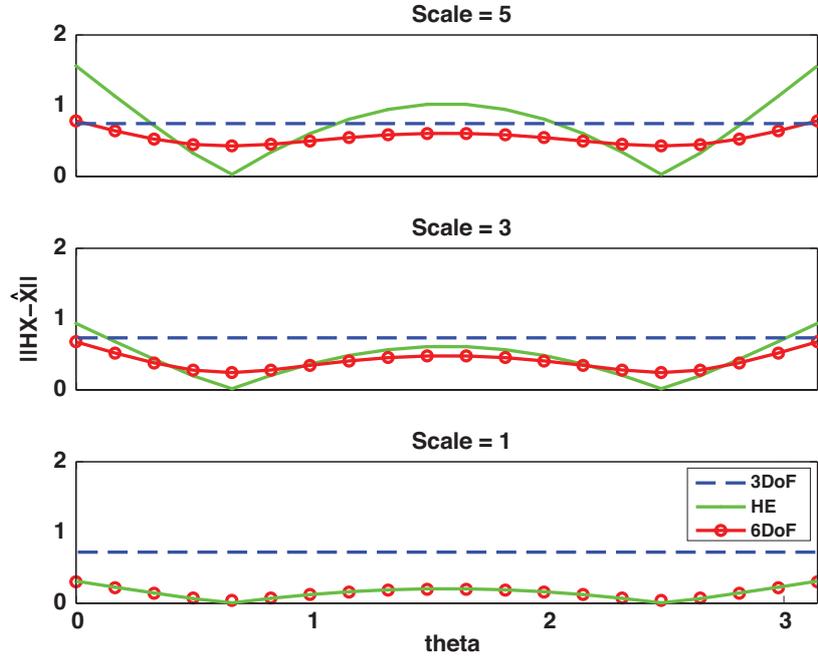


Figure 1: Comparison of the accuracy of the 3DoF method, the hand-eye Calibration (HE) method, and the 6DoF method for varying positional scale on simulated data.

minimizing over both the orientational data and the positional data. Consequently, this rotation gives a more accurate representation for 6DoF performance evaluation. It should be noted that once a rotation Ω is given, then the translation \bar{t} is calculated in the same manner for each of the three methods.

A simulation comparing all three methods is shown in Figure 1. The data were constructed by obtaining 20 equally spaced points θ_i between 0 and π . Then the positional data were constructed as

$$\begin{aligned} t_i &= [\cos(\theta_i), \sin(\theta_i), 0]^T \\ t'_i &= \Omega(\pi/3) t_i + \tau \end{aligned}$$

where \bar{t} was a randomly generated unit vector and

$$\Omega(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(x) & -\sin(x) \\ 0 & \sin(x) & \cos(x) \end{pmatrix}. \quad (10)$$

Similarly, the orientational data were constructed as

$$\begin{aligned} R_i &= \mathbf{I} \\ R'_i &= \Omega(\pi/2) \end{aligned}$$

where \mathbf{I} is the 3-dimensional identity matrix and $\Omega(\pi/2)$ is defined in Equation (10).

For each of the graphs in Figure 1 the scale for the positional data is changed. Notice as the scale increases, the fluctuations of the hand-eye calibration method increases. This is a result of the calculation of the rotation matrix Ω for the hand-eye calibration method being based solely on the orientational data. In contrast, the 3DoF method stays constant since this method is scale-invariant; while the 6DoF method, in general, stays below the 3DoF method since the 6DoF method constructs the homogeneous matrix \mathbf{H} that minimizes $\|\mathbf{H}\mathbf{X} - \mathbf{X}'\|$.

In Figure 2, the number of points θ_i between 0 and π were increased in order to compare the complexity of the algorithms. Since the calculation of the rotation for the 6DoF method is a combination of both the hand-eye calibration method and the 3DoF method, one would assume that the 6DoF method would be much slower to compute. However, the time to compute each method is approximately the same (Figure 2). This is because the order of operations for each method is the same. In order to see this point, one would just have to compare the formulation of $\bar{\mathbf{X}}\bar{\mathbf{X}}'^T$ from which the rotation matrix of each method is computed (see Section 2.1), since the rest of the algorithm for each method is identical. The operation count to compute $\bar{\mathbf{X}}\bar{\mathbf{X}}'^T$ from each method is

Method	Operation Count
3DoF	$6n^2 + 9n - 2$
Hand-Eye	$54n^2 - 3n$
6DoF	$72n^2 - 2$

which each have the same order of operations $O(n^2)$. Note that the operation counts for the 3DoF method and the 6DoF method include the operation counts from averaging and centering the positional data (see Equation (6)). In contrast, the hand-eye calibration method only includes the operation count for averaging the positional data since centering the positional data for this method is not needed.

5.3. Comparison of Methods with Real Data

A series of experiments conducted at the National Institute of Standards and Technology in November of 2009 compared the 6DoF laser tracker data (considered to be ground truth) with data collected using the eVisionFactory system (<http://www.roboticvisiontech.com/>). The eVisionFactory system calculates the rotation and translation (6DoF) by matching features of an image with features from a training image. If a specific feature is not located in an image, the system flags the corresponding rotation and translation data as being prone to errors. Hence, this data could correspond to an outlier.

Data sets from the laser tracker system and the eVisionFactory system were comprised of time-synced 6DoF data collected from a moving robot arm and a stationary object. An illustration of the setup is shown in Figure 3. Specifically, the laser tracker system collected data that consist of the active target (AT) in laser tracker (LT) coordinates (${}_{LT}\mathbf{H}_{AT}$), while the eVisionFactory system collected data that consist of the object (O) under test in camera (C) coordinates (${}_{C}\mathbf{H}_O$). Here, the active target is the reflective object from which the laser tracker system calculates the orientation and position of the target. This active target, along with the camera, are attached to the robot tool (RT) of the eVisionFactory system as can be seen in Figure 3. Therefore, a rigid homogeneous transformation

$${}_{AT}\mathbf{H}_C = {}_{AT}\mathbf{H}_{LT} \times_{LT}\mathbf{H}_{RT} \times_{RT}\mathbf{H}_C$$

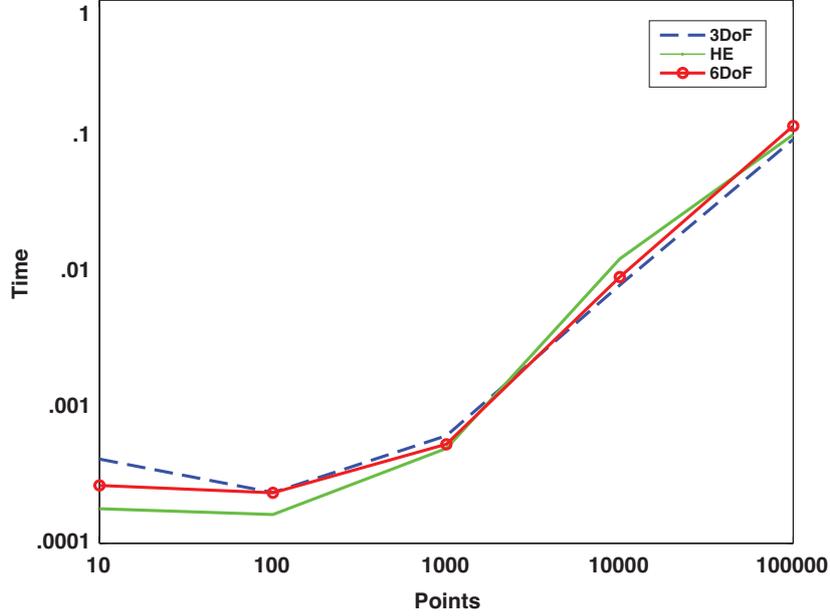


Figure 2: Comparison of the computational cost of the 3DoF method, the hand-eye Calibration (HE) method, and the 6DoF method for varying positional scale on simulated data.

from the camera to the active target can be found using external calibration techniques. Specifically, ${}_{LT}\mathbf{H}_{RT}$ can be found by rotating the robot tool from a user-defined home position amongst its three axes of rotation while recording the sequence of points traced out by the active target. After fitting circles to these sets of points, the axes of rotation of ${}_{LT}\mathbf{H}_{RT}$ are set as the normals through the center of each circle and the intersection of these normals is the origin of ${}_{LT}\mathbf{H}_{RT}$. The transformation ${}_{AT}\mathbf{H}_{LT}$ is the home position of the active target in laser tracker coordinates. Camera calibration was used to determine ${}_{RT}\mathbf{H}_C$. If the laser tracker system is reconstructed as

$${}_{LT}\mathbf{H}_C = {}_{LT}\mathbf{H}_{AT} \times {}_{AT}\mathbf{H}_C,$$

then the eVisionFactory system ${}_{O}\mathbf{H}_C$ can be compared with the laser tracker system ${}_{LT}\mathbf{H}_C$ by constructing a homogeneous matrix ${}_{LT}\mathbf{H}_O$ as outlined in this paper. Further details of the experimental setup and design can be found in [17].

An experiment was conducted where both the orientational and positional motions of the camera were adjusted [17] with results appearing in Figure 4. It should be noted that the same data was used for all methods in order to have a non-biased comparison between the three methods. The percentage of accuracy of the rotations with respect to hand-eye calibration method is nearly 100%, with respect to the 6DoF method is around 98%, and with respect to the 3DoF method is around 95%. The 6DoF method's rotational accuracy is between the 3DoF method and the hand-eye calibration method. This is a result of the 6DoF method obtaining its rotation matrix from both the orientational and positional data. In contrast, the 3DoF method obtains its rotation matrix solely from the positional data, while the hand-eye calibration method obtains its rotation matrix solely from the orientational data. The hand-eye calibration method achieves

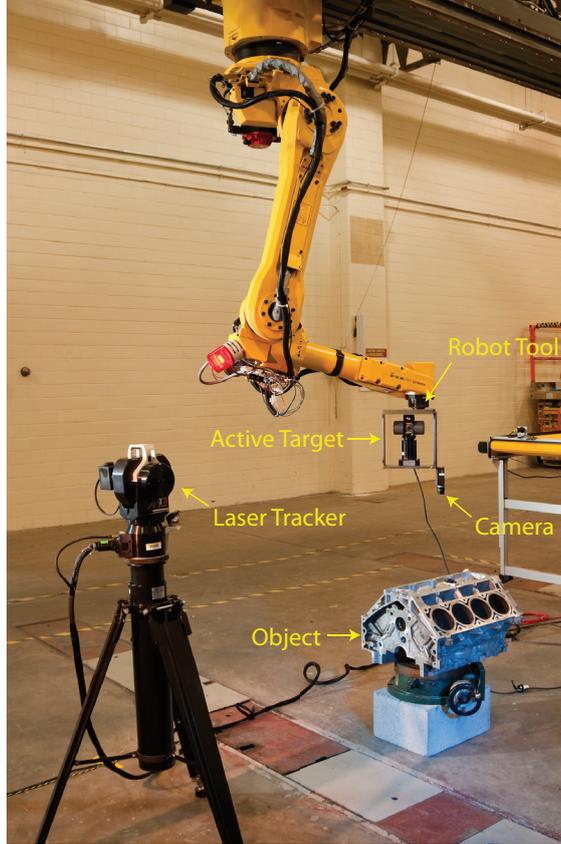


Figure 3: Experimental Setup of the eVisionFactory System.

near perfection with respect to the rotation metric since this method obtains its rotation matrix by optimizing the rotation metric; i.e. by solving $\min_{\Omega} \sum_{i=0}^{n-1} \|\Omega R_i - R'_i\|^2$. With regards to the translations, the three methods are nearly identical – all having very high accuracy. It should be noted that the homogeneous matrix computed for each method is highly dependent on the noise of the data. In this experiment, the data collected was not very noisy and thus a high level of accuracy was attained for each method. In general, this may not be the case (perhaps due to image processing and data collection) and thus more drastic differences between the three methods could be attained as is suggested in the simulated experiments of Section 5.2.

A second experiment was used to test the IQR method presented in Section 4 with results obtained from hand-calibration. Specifically, both methods constructed the homogeneous matrix ${}_{LT}\mathbf{H}_O$. It should be noted that the hand-calibration results were constructed by surveying the experimental setup and calculating the rotation and translation (hence homogeneous matrix) by hand. Consequently, the hand-calibration is prone to human error. Results are shown in Figure 5. In this figure, one can see that the IQR method locates outliers in the system (circles around the dots). These outliers match points that the eVisionFactory system acknowledges as outliers due to a feature being missing. In addition, one can see that the IQR method outperforms both the

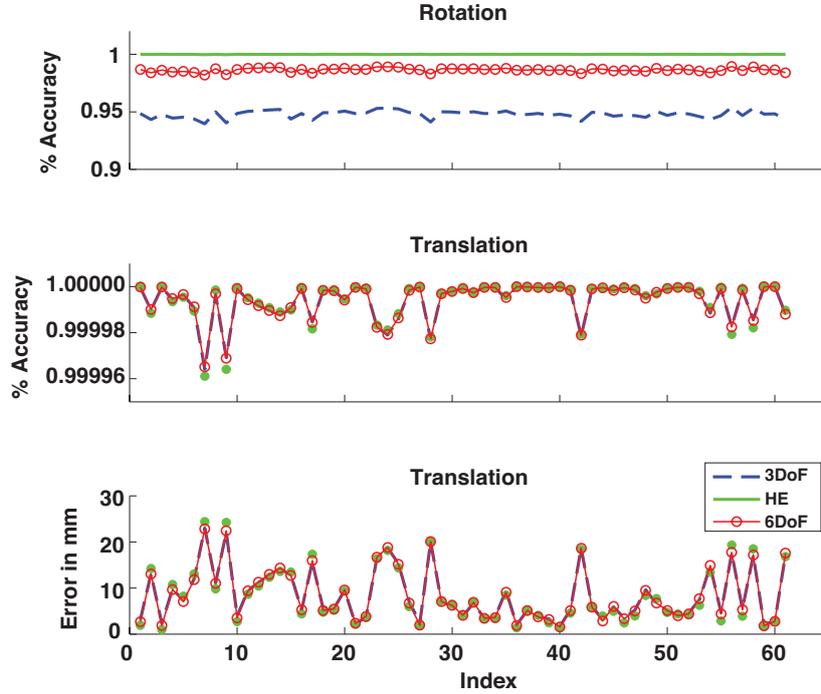


Figure 4: Comparison of the accuracy of the 3DoF method, the hand-eye calibration (HE) method, and the 6DoF method on real data obtained using the eVisionFactory system.

non-IQR and hand-calibration methods. This is due to the influence of outliers in calculating the homogeneous matrix for the non-IQR method and human error for the hand-calibration method.

6. Conclusion

In this paper, an algorithm that constructs the best homogeneous matrix \mathbf{H} that fits two sets of corresponding 6DoF data was formulated. The algorithm was tested on four experimental setups. For the first setup, the degeneracy of the 3DoF method for linear motion was explored. For the second setup, the 3DoF method, the hand-eye calibration method, and the 6DoF method were compared on simulated data, while the third setup compared the methods on real data. Finally the fourth setup tested the iterative method for identifying outliers on 6DoF data sets. For each of these experimental setups, a homogeneous matrix that represents a linear transformation from one coordinate system to the other was formulated using different methods. By comparing these different methods, the advantages of the 6DoF method was explored. Specifically, it was shown that the 6DoF method with outlier detection is an efficient and accurate method for comparing two sets of corresponding 6DoF data.

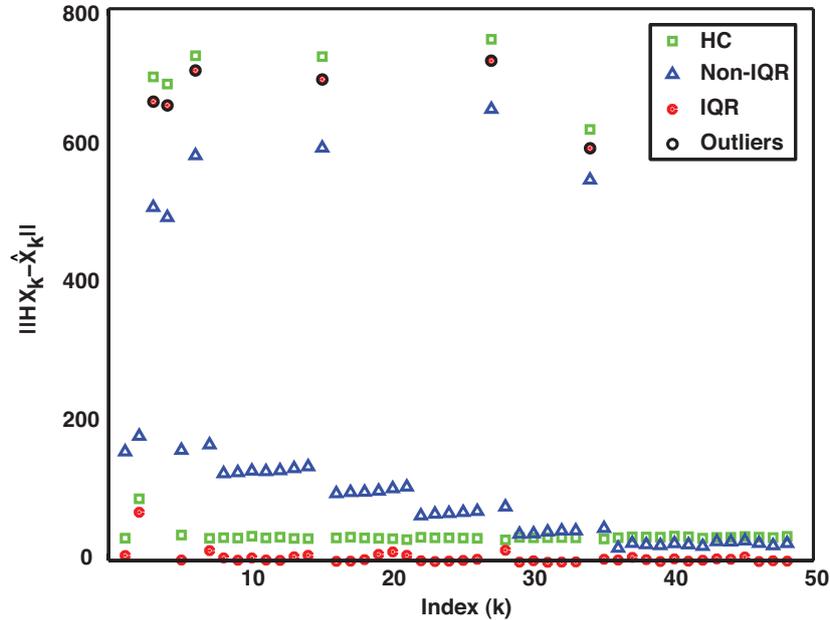


Figure 5: HC (hand-calibrated), Non-IQR, and IQR homogeneous matrix results. Notice that the IQR homogeneous matrix outperforms the hand-calibrated homogeneous matrix due to the inherent noise of the hand-calibrated results from human error. In addition, the IQR method picks out the outliers of the system.

7. Acknowledgements

The author would like to thank Roger Eastman, Tsai Hong, and Tommy Chang for all their valuable discussions and suggestions on this paper. In addition, the author would like to acknowledge Robot Vision Technologies, the National Institute of Standards and Technology, and the Purdue Robot Vision Lab for providing test data sets.

8. Role of the funding source

The author received an Interpersonal Agreement (IPA) from the National Institute of Standards and Technology (NIST). The problem discussed in this paper was formulated at NIST under the Metrology and Standards for Advanced Perception program. Data were collected and analyzed within this program with results appearing in this paper.

9. References

- [1] K. S. Arun, T. S. Huang, S. D. Blostein, Least-squares fitting of two 3-d point sets, *IEEE Trans. Pattern Anal. Mach. Intell.* 9 (5) (1987) 698–700. doi:<http://dx.doi.org/10.1109/TPAMI.1987.4767965>.
- [2] B. K. P. Horn, H. M. Hilden, S. Negahdaripour, Closed-form solution of absolute orientation using orthonormal matrices, *Journal of the Optical Society of America A* 5 (1988) 1127–1135.
- [3] B. K. P. Horn, Closed-form solution of absolute orientation using unit quaternions, *Journal of the Optical Society of America A* 4 (1987) 629–642.

- [4] M. W. Walker, L. Shao, R. A. Volz, Estimating 3-d location parameters using dual number quaternions, *CVGIP: Image Underst.* 54 (3) (1991) 358–367. doi:http://dx.doi.org/10.1016/1049-9660(91)90036-O.
- [5] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (4) (1991) 376–380. doi:http://dx.doi.org/10.1109/34.88573.
- [6] K. Kanatani, Analysis of 3-d rotation fitting, *IEEE Trans. Pattern Anal. Mach. Intell.* 16 (5) (1994) 543–549. doi:http://dx.doi.org/10.1109/34.291441.
- [7] V. M. Govindu, Lie-algebraic averaging for globally consistent motion estimation, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. 684–691.
- [8] V. M. Govindu, Consistency models for motion and calibration estimation, in: *Indian conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2000.
- [9] V. M. Govindu, Using rotational consistency for calibration estimation, in: *Conference on Information Sciences and Systems, CISS*, 2000.
- [10] F. Dornaika, R. Horaud, Simultaneous robot-world and hand-eye calibration, *IEEE Transactions on Robotics and Automation* 14 (1998) 617 – 622.
- [11] H. Zhuang, Z. S. Roth, R. Sudhakar, Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form $ax=yb$, *IEEE Transactions on Robotics and Automation* 10 (1994) 549 – 554.
- [12] K. Strobl, G. Hirzinger, Optimal hand-eye calibration, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 4647 – 4653.
- [13] C.-C. Wang, Extrinsic calibration of a vision sensor mounted on a robot, *IEEE Transactions on Robotics and Automation* 8 (1992) 161 – 175.
- [14] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.
- [15] T. Chang, T. Hong, M. Shneier, G. Holguin, J. Park, R. Eastman, Dynamic 6dof metrology for evaluating a visual servoing system, in: *Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2008, pp. 173–180.
- [16] M. Shah, T. Chang, T. Hong, R. Eastman, Mathematical metrology for evaluating a 6dof visual servoing system, in: *Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2009, pp. 182–187.
- [17] T. Chang, T. Hong, M. Shneier, M. Shah, R. Eastman, Procedure and methodology for evaluating static six-degree-of-freedom (6dof) systems, in: *Performance Metrics for Intelligent Systems (PerMIS) Workshop*, 2010.

10. Vitae



Mili Shah is an Assistant Professor at Loyola University Maryland. Dr. Shah received her B.S. in Mathematics from Emory University in 2002 and her Ph.D. in Computational and Applied Mathematics from Rice University in 2007. Dr. Shah’s research interests include computer vision, symmetry detection, eigenvalue problems, protein dynamics, and facial analysis.