

Methodology for Evaluating Static Six-Degree-of-Freedom (6DoF) Perception Systems

Tommy Chang, Tsai Hong,
Joe Falco, Michael Shneier
National Institute of Standard and Technology
Gaithersburg, Maryland
{tchang,hongt,falco,shneier}@nist.gov

Mili Shah, Roger Eastman
Loyola University Maryland
Baltimore, Maryland
{mishah, reastman}@loyola.edu

ABSTRACT

In this paper, we apply two fundamental approaches toward evaluating a static, vision based, six-degree-of-freedom (6DoF) pose determination system that measures the position and orientation of a part. The first approach uses groundtruth carefully obtained from a laser tracker and the second approach doesn't use any external groundtruth. The evaluation procedure focuses on characterizing both the system's accuracy and precision as well as the effect of object viewpoints.

For the groundtruth method, we first use a laser tracker for system calibration and then compare the calibrated output with the surveyed pose. In the method without external groundtruth, we evaluate the effect of viewpoint factors on the system's performance.

Categories and Subject Descriptors

C.4 [Performance of Systems]: Performance attributes;
B.8.2 [Performance and Reliability]: Performance Analysis and Design Aids

General Terms

Performance, Measurement, Standardization, Experimentation

Keywords

Laser Tracker, Ground Truth, 6DOF metrology, Performance Evaluation

1. INTRODUCTION

As part of the ongoing effort to standardize characterization and evaluation of 6DoF (six-degree-of-freedom) pose determination systems, we present a performance evaluation of a static, vision-based 6DoF system that measures the position and orientation of a part, also referred to as an object in this paper.

In general, performance evaluation methods can be grouped into two categories: with and without external groundtruth. We adopt the typical definition of external groundtruth as measurements obtained independently and simultaneously by an accurate system having better precision by more than one order of magnitude. Laser tracker, industrial robot arm, and computer graphics simulation are some examples of external groundtruth used for evaluating vision-based 6DoF systems [3, 14, 17].

Regardless of whether external groundtruth is employed or not, the goal is to obtain a quantitative understanding of the performance. "[The task of performance characterization] can be understood as an entirely statistical task." [16]

In this paper, we employ methods with and without external groundtruth to quantitatively determine the system's accuracy and precision and the effect of different viewpoints on its precision. For the groundtruth method, we first use a laser tracker for system calibration and then compare the calibrated result to the surveyed pose. For the method without using any external groundtruth, we estimate the system precision under various object viewpoints.

2. RELATED WORK

Here we review some common methods applied to performance characterization and evaluation of pose determination systems and algorithms [9, 3, 19, 18, 14, 13]. Other relevant work includes evaluation and characterization of ranging sensors such as LADAR (Laser Detection And Ranging) [8, 23, 1] and stereo vision [11, 17, 22], as well as algorithms including image registration [21, 5, 20, 10, 7], segmentation, and classification [2, 4]. The majority of these articles conduct experimental studies to characterize the performance under a set of controlled conditions.

In typical groundtruth methods, external groundtruth is used to compute error, which are then used to infer the unknown parameters in the error population/distribution. A statistic such as mean, standard deviation, min, and max, is computed from an error sample.

The system under test can be thought of as an estimator of the true quantity measured independently and simultaneously by the external groundtruth system. In statistics, an estimator has two properties: bias and variance. The bias measures the average *accuracy* while the variance measures the *precision* or *reliability* of the estimator [15].

One way to determine the bias of a pose determination system is to first transform data both from the groundtruth and the system under test to a common coordinate frame.

(c) 2010 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by a contractor or affiliate of the U.S. Government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

PerMIS '10, September 28-30, 2010, Baltimore, MD, USA.
Copyright © 2010 ACM 978-1-4503-0290-6-9/28/10 ...\$10.00.

Then, the bias is the averaged differences between the two transformed data sets. However, in most cases, the transformation between the two coordinate frames is not known exactly and any error in the transformation will contribute to the overall bias estimate. A robust calibration system [12] is needed to estimate the transformation.

Approaches without external groundtruth typically compute variance from the data. These variances are used to characterize and infer the effect of system’s *precision* under different conditions, or levels of the experimental factor. In [17], a dynamic 6DoF system’s performance was characterized by the standard deviation of the regression residual; although no external groundtruth was used, the motion model of the object was known. [23] and [8] characterized a LADAR system’s precision (as well as accuracy) under experimental factors including target distance, surface property, and incident angle. Statistical inference could be used to generalize and test hypotheses about the system performance.

Another approach without external groundtruth relies on an objective metric that correlates with the system’s performance. Groundtruth is still involved but only during the design and verification of the objective metric. Performance evaluation can then be done using only that objective metric alone. Examples of this approach include [22, 10, 7, 4]. In general, the objective metric is shown to vary monotonously with the amount of error, which is computed from the external groundtruth.

Unlike the previous two approaches, [9] compares the robustness of two pose estimation techniques analytically using sensitivity analysis in terms of variance amplification. [2] shows another example without the use of external groundtruth. Its idea is based on the *common agreement* metric applied to brain tissue classification: “if nine out of ten algorithms classify voxel x in subject i as white matter then one says there is a 90% chance this voxel truly is white matter.”

In this paper, we adopt the estimator approach that treats the 6DoF system under test as an estimator to the true pose. Performance of the 6DoF system can then be characterized by the estimator’s bias and variance.

3. METHODS

In this section, we describe approaches with and without external groundtruth for characterizing a commercial static, vision-based, 6DoF pose determination system.

3.1 The Static 6DoF Vision-Based System

The vision system consists of a camera mounted on a robot arm (see Figure 1). The camera used in this study has a focal length of 6mm and a resolution of 782 by 582 pixels. There are four coordinate frames involved:

1. Robot Frame: A coordinate frame located and defined at the base of the robot arm.
2. Object Frame (O): The coordinate frame associated with the object.
3. Camera Frame (C): The coordinate frame associated with the camera, which is fixed on the robot arm.
4. Source Frame (SF): A user-defined global coordinate frame relative to the robot frame. This frame is usually conveniently aligned with the object in order for the robot to perform operations relating to the object.

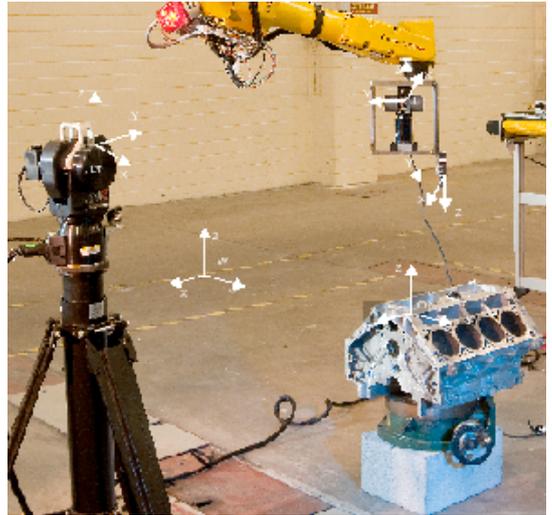


Figure 1: The vision system and the laser tracker groundtruth system.

The output from the vision-based system consists of object and camera poses for every measurement. Specifically, the outputs are three 4x4 homogeneous matrices:

- (i) ${}_{SF}\mathbf{H}_O$, transformation from the object frame to the source frame. SF is the source frame and O is the object. SF is stationary and O is also stationary.
- (ii) ${}_{SF}\mathbf{H}_C$, transformation from the camera frame to the source frame. SF is the source frame and C is the camera. In this case, SF is stationary and C is moving.
- (iii) ${}_O\mathbf{H}_C$, transformation from the camera frame to the object frame. This is obtained by combining (i) and (ii).

Although the measurement takes $\frac{1}{30}$ of a second to acquire an image, the subsequent processing time varies. The robot arm always stop moving when taking the measurement.

3.2 The GroundTruth System

The groundtruth system used in our work is a calibrated 6DoF laser tracker having a precision (two sigma) of ± 5 micrometer per meter. The laser tracker has two physical components: a portable active target that measures its own orientation and a base unit that measures the position of the active target. Together they provide the complete 6DoF pose of the active target.

There are two coordinate frames in the groundtruth system:

- Laser Tracker Frame (LT): A coordinate frame located at the base unit of the laser tracker.
- Active Target Frame (AT): The coordinate frame associated with the active target.

In our setup, we attached the active target next to the camera on the robot arm of the vision system. The output from the groundtruth system is the pose (represented by a 4x4 homogeneous matrix) of the active target:

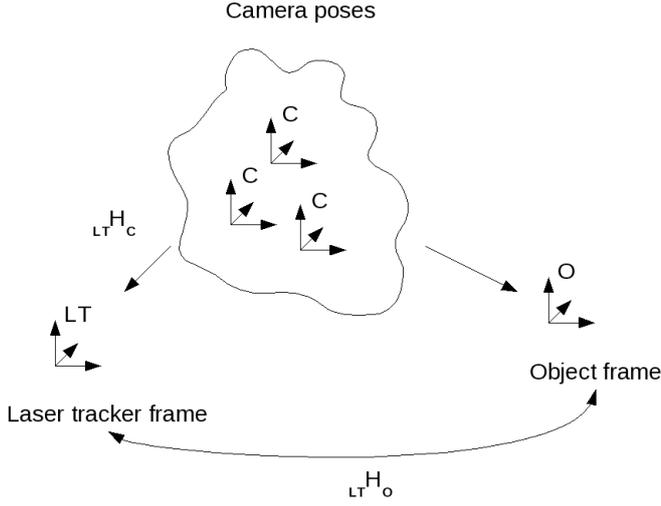


Figure 2: The dynamic camera pose is independently determined by both the vision system and the laser tracker.

- (iv) ${}_{LT}\mathbf{H}_{AT}$, transformation from the active target frame to the laser tracker frame. AT is the active target and LT is the laser tracker. Here AT moves with the robot arm and LT is stationary.

3.3 Using External GroundTruth

We are interested in using the groundtruth to estimate the bias of the vision system. A straight-forward way is to have the groundtruth system directly measure the camera frame of the vision system. This will provide the transformation between the two systems and allow camera pose and object pose, as measured by both systems, to be represented by a common coordinate frame.

However, it is extremely difficult to physically locate the camera frame. Even if we could physically locate the camera frame, there is still the issue of survey error due to operator skill and other human factors.

An equivalent approach is to numerically determine the best transformation (see Figure 2) between the set of correspondence data via optimization [13]. In order to carry out this equivalent approach, the vision system data and the laser tracker data (iv) are combined as

$${}_{LT}\mathbf{H}_C = {}_{LT}\mathbf{H}_{AT} \times {}_{AT}\mathbf{H}_C, \quad (1)$$

where ${}_{AT}\mathbf{H}_C$ was numerically determined from the robot hand-eye calibration process which included camera calibration error. Note that this transformation is constant since the active target and the camera are both rigidly mounted on the robot.

Then we construct the best-fit homogeneous matrix

$${}_{O}\hat{\mathbf{H}}_{LT} = \underset{\mathbf{H}}{\operatorname{argmin}} \|\mathbf{H} {}_{LT}\mathbf{H}_C - {}_{O}\mathbf{H}_C\|^2. \quad (2)$$

This homogeneous matrix can be constructed by first calculating the optimal rotation

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

where the full SVD of the 3×3 matrix

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

and

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

Here \mathbf{X} is the mean-adjusted data ${}_{LT}\mathbf{H}_C$ and $\hat{\mathbf{X}}$ is the mean-adjusted data ${}_{O}\mathbf{H}_C$. Once the rotation \mathbf{R} is known then the optimal translation can be calculated as

$$\mathbf{t} = \hat{\mathbf{t}} - \mathbf{R}\mathbf{t},$$

where \mathbf{t} is the mean-adjusted position data of ${}_{LT}\mathbf{H}_C$ and $\hat{\mathbf{t}}$ is the mean-adjusted position data of ${}_{O}\mathbf{H}_C$. Thus the optimal homogeneous matrix from (2) is

$${}_{O}\hat{\mathbf{H}}_{LT} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{pmatrix}.$$

Since we also can independently survey the location of the object using the laser tracker and construct ${}_{O}\mathbf{H}_{LT}$, we can compare the results of the best-fit homogeneous matrix

$${}_{O}\hat{\mathbf{H}}_{LT}$$

with the groundtruth ${}_{O}\mathbf{H}_{LT}$. The difference between the best-fit matrix and the surveyed groundtruth matrix include the system bias error and the groundtruth measurement error.

3.4 Without using External GroundTruth

In a way, an implicit groundtruth is present in the static scenario. This implicit groundtruth is embedded into the experimental setup by having the object remained stationary. We knew the object did not move, therefore, we don't need a real external groundtruth system to measure its pose, which is just a constant by design.

3.4.1 Effect of object Viewpoints

We investigate the effect of different object viewpoints on the variance of the vision system. [9] studied two pose determination algorithms and showed, analytically and experimentally, that viewpoint has an effect on pose stability. In our study, instead of varying the object pose, experiments were setup to vary the camera pose. The effect of varying the camera pose is the same as varying the object pose because in both cases the camera produces the same object image. In addition, we gained an implicit groundtruth from the fact that the object remained stationary.

Four independent experiments are conducted. In each experiment, the pose of a static object is to be compared under three different measurement conditions (treatments): A, B, C. The experiments are described below:

- Exp0: scale factor

A scale of 1.0 means that the entire object, regardless of its orientation, occupies the image as much as possible. A scale of 0.5 means that at most 2 objects can be seen simultaneously in the image, regardless of their orientations. To compute the actual scale, we first determine the minimum bounding sphere of the object and then measure the distance from camera to the object. Using the pin-hole camera model, we can then determine the scale by

$$\text{scale} = \frac{\text{dia}}{2 \times \text{dist} \times \tan\left(\frac{\text{fov}}{2}\right)},$$

where dia is the diameter of the minimum bounding sphere, dist is the distance between the camera and the object, and fov is the camera’s vertical field-of-view angle.

- cond A: object seen at low-scale (computed scale = 0.37)
- cond B: object seen at mid-scale (computed scale = 0.55)
- cond C: object seen at full-scale (computed scale = 1.03)
- Exp1: position factor
 - cond A: object seen near the image border (actual camera shift = 100 mm)
 - cond B: object seen midway between center of FoV (field-of-view) and the image border (actual camera shift = 50 mm)
 - cond C: object seen at the center of FoV (0 shift)
- Exp2: azimuth factor

Under this factor, the object is always centered in the image as the camera rotates about its optical axis.

 - cond A: object seen rotated 40 degree
 - cond B: object seen rotated 20 degree
 - cond C: object seen up-right
- Exp3: polar factor

Under this factor, the object is always centered in the image as the camera rotates around the object.

 - cond A: object seen rotated 25 degree on its side
 - cond B: object seen rotated 15 degree on its side
 - cond C: object seen up-right

Except experiment Exp0, the object scale is fixed at 0.55 throughout the experiments.

4. RESULTS

This section describes the dataset obtained and their preliminary analysis.

4.1 Data without External Groundtruth

In each of the four viewpoint experiments, 10 runs per treatment were carried out, resulting a total of 30 runs per experiment. More runs could be used, but 10 were chosen to establish an initial preliminary study. No groundtruth data was collected in these viewpoint experiments.

We used the completely randomized design (CRD) paradigm [15] in our viewpoint experiments and identified time and robot repeatability as two nuisance factors that we have no control over. However, we did not randomize the order of runs (as to neutralize the possible timing and robot effect) for two reasons:

1. The condition/treatment order can not be changed. The commercial vision system always perform conditions A, B, C, A, B, C, ... in that cyclic order. The provided commercial software does not have an option to change the run order.

2. The robot arm was found¹ to have deterministic repeatability after warming up for 20 minutes. Therefore, the robot’s performance (repeatability as specified by the manufacture) does not change with time. However, it was noted that the robot’s repeatability depends on the motion as well as the initial pose at the time the motion command was issued. As a result, we always move the robot from a fixed initial pose and set the robot speed to low (as to minimize structural vibration caused by robot motion).

4.2 Data with External Groundtruth

Additionally, four data sets were collected together with the groundtruth. Groundtruth was obtained by matching the vision data with the corresponding laser tracker data. Since the clocks were synchronized and timestamps recorded, we can match data by their timestamp. For each data set, the camera height was measured to about 457mm.

- For the first data set, we adjusted only the rotational motion of the camera. Specifically, we rotated the camera about the rotational axes R_x and R_y from ± 15 degrees in increments of 5 degrees, and R_z from ± 10 degrees in increments of 5 degrees. (Total of $7 \times 7 \times 5 = 245$ data; 61 have all image features detected.)
- For the second data set, we adjusted only the translational motion of the camera. Specifically, we moved the camera in the x and y directions from ± 150 mm in increments of 50mm. (Total of $7 \times 7 = 49$ data; 37 have all image features detected)
- For the third data set, we adjusted both the rotational and translational motion of the camera. Specifically, we rotated the camera about the rotational axes R_y and R_z from ± 5 degrees in increments of 5 degrees, and moved the camera in the x direction from ± 150 mm in increments of 50mm. (Total of $3 \times 3 \times 7 = 63$ data; 54 have all image features detected)
- For the fourth data set, we adjusted both the rotational and translational motion of the camera. Specifically, we rotated the camera about the rotational axes R_y and R_z from ± 5 degrees in increments of 5 degrees, and moved the camera in the y direction from ± 150 mm in increments of 50mm. (Total of $3 \times 3 \times 7 = 63$ data, 41 have all image feature detected)

4.3 External GroundTruth Result

Table 1 summarizes error between the optimal homogeneous matrix ${}_{O}\widehat{\mathbf{H}}_{LT}$ computed for each of the four data sets with the survey from the laser tracker ${}_{O}\mathbf{H}_{LT}$. The error shown in this table is the combined effect of survey error and the system bias. It should be noted that certain data points were ignored in the calculation of the best homogeneous matrix ${}_{O}\widehat{\mathbf{H}}_{LT}$. These points correspond to positions where all the image features could not be located by the vision system. Outliers were also removed in the construction of the optimal homogeneous matrix. These outliers were constructed using a statistical tool that identifies points as outliers if they lie outside of one and half times the interquartile range.

¹We used the IS09283 robot performance standard metric and protocol describe in [6].

Table 1: Estimates of system bias

| | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
|----------|--------|--------|--------|----------|----------|----------|
| DataSet1 | 8.696 | 2.868 | 47.191 | 0.0328 | 0.0038 | 0.2435 |
| DataSet2 | 9.272 | 3.476 | 39.755 | 0.0767 | 0.0336 | 0.1544 |
| DataSet3 | 8.946 | 6.760 | 37.092 | 0.0572 | 0.0069 | 0.0522 |
| DataSet4 | 9.569 | 4.687 | 40.551 | 0.1160 | 0.0074 | 0.1967 |

Table 1 indicates a large combined error in positional measurements. The Z-component has the largest error, and Y-component has the least error. This is not surprising, given that the image appearance typically changes only slightly as the height of the camera changes, and the pose detection algorithm depends on only a single camera view.

Tables 2 to 6 summarize the error residual between the groundtruth data and the corresponding transformed vision data in the same coordinate. Before finding the best-fit transformation, some outlier points were removed from the groundtruth data. The points arose because the laser tracker has a problem tracking jerks in the robot’s motion. Data for a short time following a jerk are incorrect. Unfortunately, this problem was only discovered after the data had been collected, and the sampling did not wait long enough for the system to settle after the camera reached its destination before collecting data. The pairs of points corresponding to these measurements were omitted from the best-fit calculations. However, the errors in the tables 2 to 6 are computed from all the collected data points that the vision system was able to detect all image features.

The requirements for most applications for which the 6DoF system were developed depends on repeatability rather than accuracy, but even here, the variation can be large. For example, in data set 4 (summarized in Table 5 and Figure 6) the system erroneously matches features. Note, however, the data set actually produced the best results when the bad matches are omitted.

Results from the four data sets are shown in Figure 3 to 6. Points were selected as outliers if they lay outside one and half times the inter-quartile range. It should be noted that there are close fits for most of the points but there are few points with a large error. These points correspond to positions where the vision system indicated a good match to the data (all features were detected) but two or more of the detected features matched to the wrong model feature.

Overall, the user would have to decide if the system was repeatable enough for a particular application. The performance data provide the necessary information to do so. The mean and standard deviation of the error, together with the maximum errors, can be compared with the tolerances of the application. They can also be used in process control; if a measured part location lies, for example, more than two standard deviations from mean, it likely indicates either a bad part or an erroneous match between the part and the model. The vision system could attempt to reacquire the part and if it failed again, the part could be rejected.

4.4 Result without GroundTruth

Tables 7 to 10 summarize the system variance (computed as standard deviation) among the three conditions in each of the four viewpoint experiments (see Section 3.4.1). Originally, the experiment was setup to answer the specific question: “Do pose solutions from the vision system differ significantly under different object viewpoint?” In statistics,

Table 2: Data Set 1 — Rotation

| 61 poses with all image features detected | | | | | | |
|-------------------------------------------|--------|--------|--------|----------|----------|----------|
| | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
| Mean | 1.709 | 1.628 | 0.846 | 0.1824 | 0.0911 | 0.1896 |
| Median | 1.542 | 1.300 | 0.723 | 0.1621 | 0.0854 | 0.1504 |
| Std Dev | 1.109 | 1.437 | 0.537 | 0.1241 | 0.0582 | 0.1763 |
| Min | 0.043 | 0.001 | 0.008 | 0.0072 | 0.0020 | 0.0029 |
| Max | 4.391 | 5.113 | 2.551 | 0.5252 | 0.2273 | 0.9390 |

Table 3: Data Set 2 — Translation

| 37 poses with all image features detected | | | | | | |
|-------------------------------------------|--------|--------|--------|----------|----------|----------|
| | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
| Mean | 1.736 | 5.848 | 1.100 | 0.1472 | 0.0332 | 0.5146 |
| Median | 1.535 | 4.397 | 0.866 | 0.1139 | 0.0234 | 0.3579 |
| Std Dev | 1.506 | 4.367 | 0.847 | 0.1316 | 0.0297 | 0.3870 |
| Min | 0.024 | 0.036 | 0.004 | 0.0023 | 0.0002 | 0.0402 |
| Max | 6.879 | 16.655 | 3.461 | 0.5660 | 0.1199 | 1.4359 |

Table 4: Data Set 3 — Rotation and Translation

| 54 poses with all image features detected | | | | | | |
|-------------------------------------------|--------|--------|--------|----------|----------|----------|
| | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
| Mean | 2.454 | 8.559 | 0.730 | 0.1934 | 0.0764 | 0.7082 |
| Median | 1.678 | 5.354 | 0.676 | 0.1137 | 0.0753 | 0.4699 |
| Std Dev | 3.503 | 8.453 | 0.579 | 0.2799 | 0.0401 | 0.6907 |
| Min | 0.018 | 0.174 | 0.039 | 0.0003 | 0.0025 | 0.0106 |
| Max | 24.036 | 42.317 | 3.464 | 1.9085 | 0.1664 | 3.5107 |

Table 5: Data Set 4 — Rotation and Translation

| 41 poses with all image features detected | | | | | | |
|-------------------------------------------|---------|---------|---------|----------|----------|----------|
| | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
| Mean | 20.815 | 8.260 | 46.677 | 21.9294 | 2.0078 | 3.8628 |
| Median | 1.226 | 2.464 | 0.559 | 0.0963 | 0.0872 | 0.2285 |
| Std Dev | 70.215 | 21.247 | 165.772 | 78.6847 | 8.4709 | 12.9773 |
| Min | 0.017 | 0.360 | 0.057 | 0.0015 | 0.0048 | 0.0162 |
| Max | 294.918 | 109.988 | 645.663 | 307.4627 | 51.4941 | 57.3316 |

Table 6: Combined Data

| Combined Data (total of 193 poses) | | | | | | |
|------------------------------------|---------|---------|---------|----------|----------|----------|
| | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
| Mean | 5.941 | 5.778 | 10.498 | 4.7513 | 0.4790 | 1.1699 |
| Median | 1.461 | 2.724 | 0.686 | 0.1217 | 0.0677 | 0.2657 |
| Std Dev | 32.864 | 11.232 | 77.567 | 36.8178 | 3.9272 | 6.0726 |
| Min | 0.017 | 0.001 | 0.004 | 0.0003 | 0.0002 | 0.0029 |
| Max | 294.918 | 109.988 | 645.663 | 307.4627 | 51.4941 | 57.3316 |

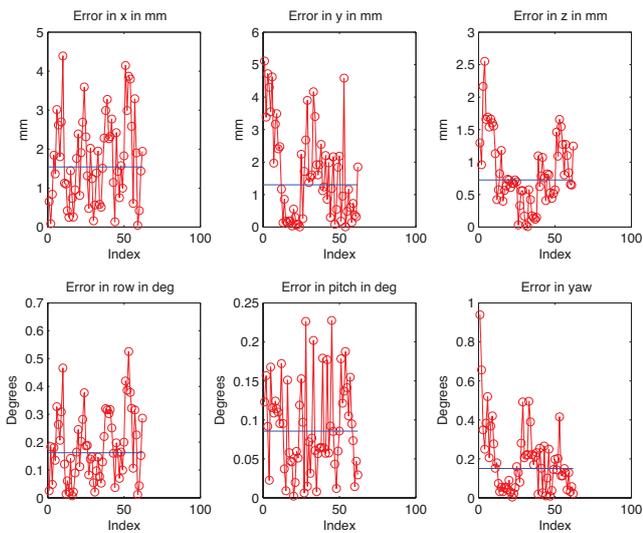


Figure 3: Data Set 1 — Rotation Only

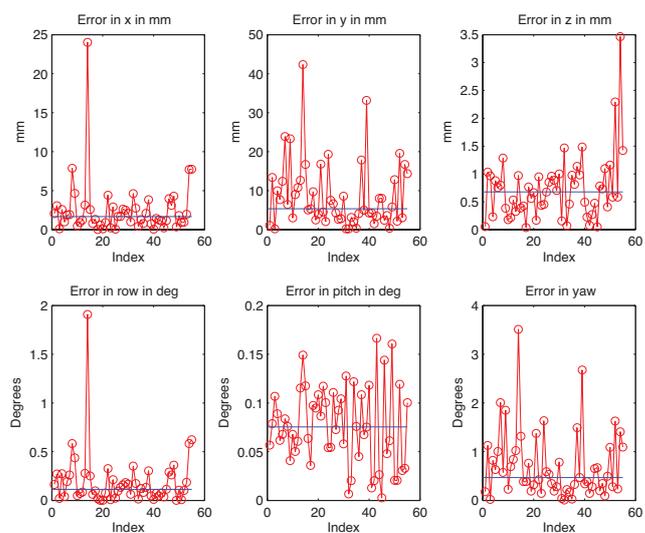


Figure 5: Data Set 3 — Rotation and X Translation

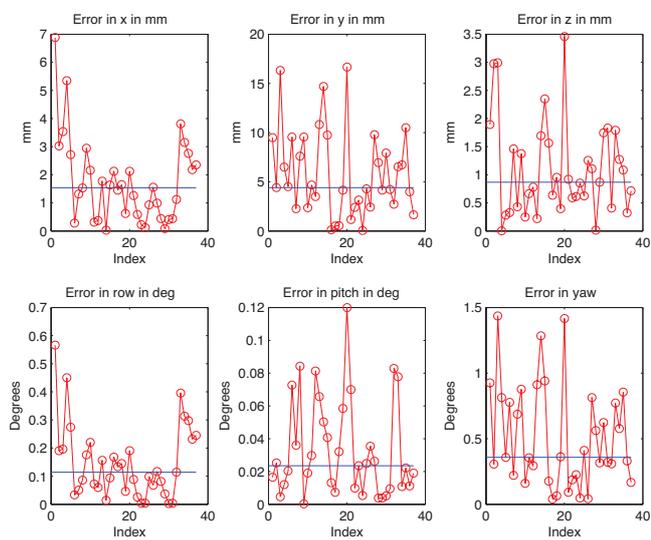


Figure 4: Data Set 2 — Translation Only

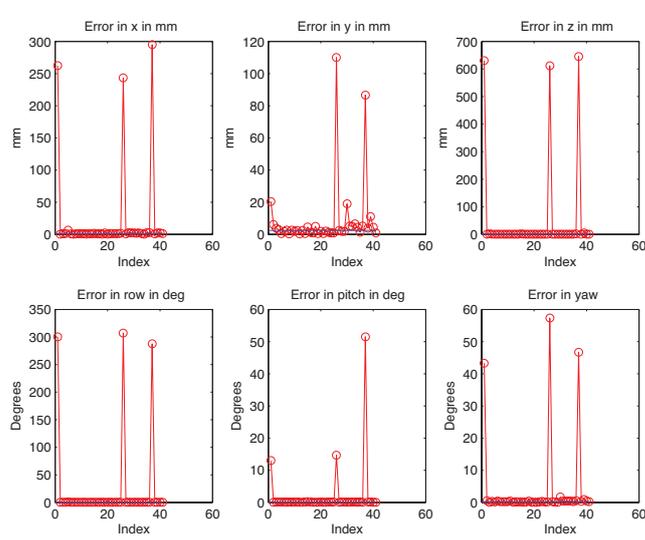


Figure 6: Data Set 4 — Rotation and Y Translation

Table 7: Exp0 — Scale Factor

| Std Dev | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
|---------|--------|--------|--------|----------|----------|----------|
| CondA | 0.038 | 0.025 | 0.070 | 0.0344 | 0.0643 | 0.0096 |
| CondB | 0.046 | 0.018 | 0.078 | 0.0409 | 0.0521 | 0.0051 |
| CondC | 0.052 | 0.137 | 1.043 | 1.8300 | 0.3541 | 0.0622 |

Table 8: Exp1 — Position Factor

| Std Dev | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
|---------|--------|--------|--------|----------|----------|----------|
| CondA | 0.057 | 0.021 | 0.040 | 0.1145 | 0.1445 | 0.0198 |
| CondB | 0.072 | 0.026 | 0.071 | 0.1680 | 0.1532 | 0.0164 |
| CondC | 0.087 | 0.017 | 0.073 | 0.1080 | 0.0585 | 0.0048 |

this type of question is commonly answered by testing the homogeneity hypothesis [15].

In our experiment, we collected just 10 measurements per each of the three conditions. Without a priori knowledge about the underlying population distributions, small samples can not be justified for use in testing the homogeneity hypothesis. Since we don't know the underlying populations, one idea is to use non-parametric approaches, which make few assumptions about the population distribution. One non-parametric approach we considered the Kruskal-Wallis test [15], which assumes the populations all have the same shape. However, since our sample size was small this assumption could not be verified.

Instead of testing the homogeneity hypothesis, another approach to answer our original question is to use statistical methods that directly compare distributions. Such methods include χ^2 goodness-of-fit test, Kolmogorov-Smirnov goodness-of-fit test, and others. However, due to the small sample size and the uncertainty about our samples being representative of their populations, we did not pursue any of these methods. Nevertheless, the collected data provide some useful insights:

- From the Exp0 (scale factor) data summarized in Table 7, we observed that condition C has the largest variance compared to the two other conditions. With the exception of the X component, the variances are at least 5 times larger in condition C. Since condition C corresponds to the smallest image scale, it implies that a sudden degradation in system precision can be expected when the object in the image gets smaller than a threshold.
- With the scale set at 0.55, Tables 8 to 10 show a combined system precision that is better than 0.7° for orientation and 0.3 mm for position. The combined conditions encompass a viewpoint coverage of up to 100 mm shift in X position, 40 degrees in azimuth angle, and 25 degrees in polar angle.
- We noted that the physical dimension is a function of camera lens. By changing the camera lens, the vision system can be adapt to larger or smaller object. It is reasonable to think that both the accuracy and precision of the system will improve if a higher resolution camera is used.

5. CONCLUSION

Table 9: Exp2 — Azimuth Factor

| Std Dev | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
|---------|--------|--------|--------|----------|----------|----------|
| CondA | 0.010 | 0.029 | 0.032 | 0.0163 | 0.0485 | 0.0206 |
| CondB | 0.022 | 0.038 | 0.035 | 0.0245 | 0.0864 | 0.0123 |
| CondC | 0.024 | 0.026 | 0.069 | 0.0458 | 0.1092 | 0.0235 |

Table 10: Exp3 — Polar Factor

| Std Dev | X (mm) | Y (mm) | Z (mm) | Rx (deg) | Ry (deg) | Rz (deg) |
|---------|--------|--------|--------|----------|----------|----------|
| CondA | 0.097 | 0.051 | 0.294 | 0.2112 | 0.2189 | 0.0143 |
| CondB | 0.040 | 0.019 | 0.116 | 0.1191 | 0.6834 | 0.0756 |
| CondC | 0.017 | 0.032 | 0.037 | 0.1730 | 0.0387 | 0.0048 |

We described and applied two common approaches for evaluating and characterizing the performance of 6DOF perception systems. External groundtruth is necessary to evaluate system accuracy in terms of its bias. System precision, in terms of its repeatability, can be evaluated with or without an external groundtruth. In both cases, the result characterizes the system under the condition in which it was operated.

For evaluating 6DoF systems, the use of an external groundtruth system is essential when the pose is dynamic or static. The mean and standard deviation of the errors, together with the maximum errors, can be compared with the tolerance of the user's requirements in their application. The user's requirements will decide if the system is repeatable enough for a particular application. However, if the external groundtruth is not available, then system uncertainty using the variance in the data may be an alternative approach for estimating the system precision.

In the case of no external groundtruth, our approach was to test whether the homogeneity hypothesis could be used to answer the question: "Do pose solutions from the vision system differ significantly under different object viewpoint?" Unfortunately, given the insufficiency of data we collected, the lack of a priori knowledge about the underlying population distributions, and the doubt about samples being representative, we were unable to justify using and applying the homogeneity hypothesis.

What performance factors to study depends on the intended application of the system. In our case, we used object viewpoint as an example. The users of 6DoF pose systems may be interested in other factors such as environmental lighting, object type, object pose, object motion, operator skill, etc.

Our long-term goal is to assist in developing a standard for performance evaluation of dynamic 6DoF measurement systems. This standard will specify quantitative, reproducible test methods to evaluate the robustness, accuracy, repeatability and other performance characteristics of dynamic 6DoF systems. The standard will also assist in the development of new applications of automation by enabling end-users to directly compare 6DoF systems as well as reducing the time spent on system evaluation, adoption and integration.

6. ACKNOWLEDGMENTS

We thank Dennis Murphy, Clifton Miskov, and Remus Boca for their technical supports in carrying out this study.

7. REFERENCES

- [1] D. Anderson, H. Herman, and A. Kelly. Experimental characterization of commercial flash lidar devices. In *International Conference of Sensing and Technology*, November 2005.
- [2] S. Bouix, M. Martin-Fernandez, L. Ungar, M. N. M.-S. Koo, R. W. McCarley, and M. E. Shenton. On evaluating brain tissue classifiers without a ground truth. *NeuroImage*, 36:1207–1224, 2007.
- [3] T. Chang, T. Hong, M. Shneier, G. Holguin, J. Park, and R. D. Eastman. Dynamic 6dof metrology for evaluating a visual servoing system. In *PerMIS '08: Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pages 173–180, New York, NY, USA, 2008. ACM.
- [4] C. E. Erdem, A. M. Tekalp, and B. Sankur. Metrics for performance evaluation of video object segmentation and tracking without ground-truth. *International Conference On Image Processing*, 2001.
- [5] J. Fitzpatrick and J. West. A blinded evaluation and comparison of image registration methods. In *IEEE Workshop on Empirical Evaluation Methods in Computer Vision*, 1998.
- [6] ISO 9283:1998. *Manipulating industrial robots — Performance criteria and related test methods. Second edition*. ISO, Geneva, Switzerland.
- [7] J. Kybic. Fast no ground truth image registration accuracy evaluation: Comparison of bootstrap and hessian approaches. In *ISBI*, pages 792–795, 2008.
- [8] A. M. Lytle, S. Szabo, G. Cheok, K. Saidi, and R. Norcross. Performance evaluation of a 3d imaging system for vehicle safety. In *Unmanned Systems Technology IX. Proceedings of the SPIE, Volume 6561*, 2007.
- [9] C. B. Madsen. A comparative study of the robustness of two pose estimation techniques. *Machine Vision and Applications*, 9(5-6):291–303, 1997.
- [10] R. Schestowitz, C. J. Twining, T. Cootes, V. Petrovic, B. Crum, and C. J. Taylor. Non-rigid registration assessment without ground truth. In *Medical Image Understanding and Analysis*, volume 2, pages 151–155, 2006.
- [11] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 519–528, Washington, DC, USA, 2006. IEEE Computer Society.
- [12] M. Shah, T. Chang, and T. Hong. Mathematical metrology for evaluation a 6dof visual servoing system. In *PerMIS '09: Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems*, 2009.
- [13] M. I. Shah. Six degree of freedom point correspondences. *Technical Report, Loyola University Maryland, TR2009_02.*, 2009.
- [14] F. Smit and R. van Liere. A framework for performance evaluation of model-based optical trackers. *EGVE Symposium*, 6, 2008.
- [15] A. C. Tamhane and D. D. Dunlop. *Statistics and Data Analysis from Elementary to Intermediate*. Prentice Hall, Upper Saddle River, NJ 07458, 2000.
- [16] N. A. Thacker, A. F. Clark, J. L. Barron, J. R. Beveridge, P. Courtney, W. R. Crum, V. Ramesh, and C. Clark. Performance characterization in computer vision: A guide to best practices. *Computer Vision and Image Understanding*, 109(3):305–334, 2008.
- [17] M. Tonko and H.-H. Nagel. Model-based stereo-tracking of non-polyhedral objects for automatic disassembly experiments. *International Journal of Computer Vision*, 37(1):99–118, 2000.
- [18] R. van Liere and A. van Rhijn. An experimental comparison of three optical trackers for model based pose determination in virtual reality. In S. Coquillart and M. Göbel, editors, *10th Eurographics Symposium on Virtual Environments*, pages 25–34, Grenoble, France, 2004. Eurographics Association.
- [19] F. Vikstén, P.-E. Forssén, B. Johansson, and A. Moe. Comparison of local image descriptors for full 6 degree-of-freedom pose estimation. In *IEEE International Conference on Robotics and Automation*, Kobe, Japan, May 2009. IEEE, IEEE Robotics and Automation Society.
- [20] J. B. West and J. M. Fitzpatrick. The distribution of target registration error in rigid-body, point-based registration. In *IPMI '99: Proceedings of the 16th International Conference on Information Processing in Medical Imaging*, pages 460–465, London, UK, 1999. Springer-Verlag.
- [21] J. B. West, J. M. Fitzpatrick, M. Y. Wang, B. M. Dawant, C. R. M. Jr., R. M. Kessler, and R. J. Maciunas. Retrospective intermodality registration techniques for images of the head: Surface-based versus volume-based. *IEEE Trans. Med. Imaging*, 18(2):144–150, 1999.
- [22] Q. Yang, R. M. Steele, D. Nister, and C. Jaynes. Learning the probability of correspondences without ground truth. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1140–1147, Washington, DC, USA, 2005. IEEE Computer Society.
- [23] C. Ye and J. Borenstein. Characterization of a 2-d laser scanner for mobile robot obstacle negotiation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2512–2518, 2002.