Ground Truth for Evaluating 6 Degrees of Freedom Pose Estimation Systems

Jeremy A. Marvel
National Institute of Standards and Technology
100 Bureau Drive, MS 8230
Gaithersburg, MD 20899
jeremy.marvel@nist.gov

Joe Falco
National Institute of Standards and Technology
100 Bureau Drive, MS 8230
Gaithersburg, MD 20899
joseph.falco@nist.gov

Tsai Hong
National Institute of Standards and Technology
100 Bureau Drive, MS 8230
Gaithersburg, MD 20899
tsai.hong@nist.gov

ABSTRACT

Systems developed to estimate poses of objects in 6 degrees of freedom (6DOF) Cartesian space (X, Y, and Z coordinates plus roll, pitch, and yaw) are reliant on the vendors' own processes to determine performance and measurement accuracy. These practices are not yet standardized, and are rarely reported by the vendors in sufficient detail to enable users and integrators to recreate the process. Efforts must therefore be made to enable the documented and, more importantly, independently repeatable evaluation of such systems using standardized processes, fixtures, and artifacts. In this paper, we describe three 6DOF ground truth systems utilized at the National Institute of Standards and Technology (NIST): a laser-tracker-based system for pose measurement, an aluminum fixture-based system that can be used to set the pose of artifacts, and a modular, medium-density fiberboard (MDF) fixture system. Descriptions, characterizations, and measured accuracies of these systems are provided for reference.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Performance Attributes; B.8.2 [**Performance and Reliability**]: Performance Analysis and Design Aids; G.1.6 [**Optimization**]: I.5.4 [**Applications**]: Computer Vision

General Terms

Measurement, Documentation, Performance, Experimentation, Standardization, Verification

Keywords

Ground Truth, 6DOF Metrology, Laser Tracker, Fixtures

1. INTRODUCTION

The usefulness of novel 6DOF pose estimation systems is restricted only by the accuracy with which it can measure objects in the world space. Reporting this is relatively simple, but the initial evaluation and subsequent validation of the reported values are extensive processes requiring the appropriate metrics and either a reference standard system (i.e., an external ground truth) against which the accuracy of a system under test can be measured, or a methodology of computing variances in the data to infer a given system's precision under different operational conditions. The utilization of ground truths is a fundamental aspect of measurement science, and provides a basement of comparison for the estimated quantities measured independently and simultaneously by the system under test.

A fundamental limitation of ground truth utilization, however, lies in the difficulty in obtaining the ground truth, itself. Establishing a measurement system as a ground truth requires extensive efforts and measurement tools in validating its accuracy. As a general rule, the ground truth system must be at least an order of magnitude more accurate than the system under test. The tools required to assess the accuracy of potential ground truths are prohibitively expensive, and must conform to set traceability standards, themselves, in their establishment as ground truths.

In this paper we discuss the evolution of NIST-developed ground truth systems in efforts to make 6DOF metrology evaluation more accessible and expandable. Three different systems are presented, and their accuracies and measurement uncertainties are provided. The issues addressed in this report focus on the development and validation of ground truth systems, and are discussed in an effort to provide examples of the establishment of new ground truths.

2. RELATED WORK

Although 3D pose estimation systems may be evaluated sans ground truth [1], the utilization of an external ground truth is typical for measurement systems for the computation of errors in pose estimations. These errors are then evaluated to infer statistical distribution (mean, standard deviation, and error trends) of the bias and variance of the environmental parameter space for the sensor under test [2].

The ground truths may be either sensor- or artifact-based, and are expected to be at least an order of magnitude more accurate than the system under test. Sensor-based ground truths—where the pose of an object is based on the measured outputs of a system with known accuracy and precision—are traditionally flexible and modular in nature, but require a robust calibration system [3] to establish a common coordinate frame between the ground truth and sensor under test. While many ground truth systems employ some form of fiducial attached to the surface of an evaluated target in order to enable precision metrology (e.g., laser-tracked active targets [4] and camera-tracked active [5] and passive [6]

This paper is authored by employees of the United States Government and is in the public domain. PerMIS'12, March 20-22, 2012, College Park, MD, USA. ACM 978-1-4503-1126-7-3/22/12

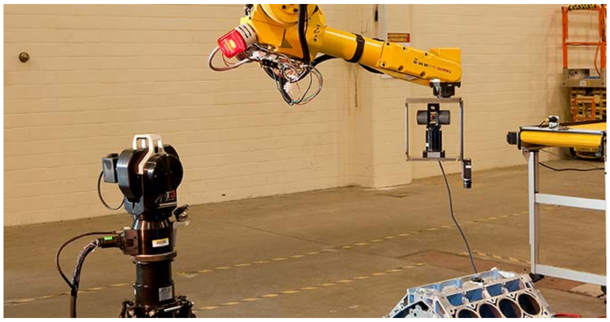

Figure 1. The laser tracker ground truth system (left) and active target attached to an industrial robot arm (right).

Table 1. Uncertainties (Standard Deviations) of Position (X, Y, and Z) and Rotation (Roll, Pitch, and Yaw) Measurements of the Laser Tracker System

	X	Y	Z
Uncertainty (mm)	0.0018	0.0014	0.0021
	Roll	**Pitch**	**Yaw**
Uncertainty (degrees)	0.0007	0.0001	0.0012

targets), not all evaluation systems are compatible with them. Such artifacts may inadvertently change the surface properties of the target, and thus interfere with the performance of certain shape- and feature-based pose estimation systems.

Artifact-based ground truths are based on either fixtured components with associated *a priori* knowledge of transformations and pose uncertainty, or known distributions of features on a specific truthing object. In contrast with the sensor-based ground truths, artifact-based ground truths are typically easier to use in evaluations, are generally more readily repeatable, and are more affordable and accessible to a variety of researchers. For instance, in [7] a simplified cardboard artifact was rigidly affixed to a rotational base for a single DOF in pose variance. The rotational base had position sensor to read orientation angle around a pivot point. Further, in [8], rigid automotive engine components were used for validation of their 3D pose estimation system using feature-based tracking of various component assembly points relative to one another.

Artifact-based ground truths involving physical objects, however, are subject to measurement uncertainties in pose and adherence to construction tolerances, both of which necessarily introduce some error in establishing the ground truth. As such, an alternative artifact-based approach utilizes synthetic data for test and evaluation of pose estimation systems. For instance, [9] utilized computer-generated images of geometric primitives with associated CAD models to evaluate a proposed single-camera 3D pose estimation system, while [10] used simulated 3D point cloud data and robot pose information to validate a 6DOF localization methodology using polygonal indoor maps.

3. LASER TRACKER

The laser tracker system, shown with its active target in a testing configuration in Figure 1, has been utilized as a high-precision ground truth for 3D measurements at NIST since 2008 (e.g., [1, 3, 11]). It has been used to truth component positions of manufacturing and construction systems when tolerance accuracies are unknown or unreliable. The laser tracker boasts high measurement accuracy, but at the expense of monetary cost. The full cost of the system utilized at NIST is approximately $150 000.

3.1 Measurement Configuration

The laser tracker configuration utilized for 6DOF pose measurement has two physical components: a portable active target that measures its own orientation using a motorized receiver and a level sensor, and a base laser unit that measures the position of the active target [4]. Together, they provide the complete 6DOF pose of the active target. The active target can only be used for measuring static 6DOF poses with a precision of ±3 arc-seconds in angle (±0.0008 degrees), and a combined positional accuracy of 15 µm average error with uncertainty of 10 µm at 2.0 m. The active target, which is either attached to or substituted in lieu of the object to be truthed, requires a direct line of sight with the laser unit's beam, and thus only one object can be measured at a time.

3.2 Measured Accuracy

The measurement accuracy of the laser tracker system mentioned earlier was specified by the manufacturer. These accuracy specifications were validated according to the process laid out in the ASME B89.4.19-2006 standard (*Performance Evaluation of Laser-Based Spherical Coordinate Measurement Systems*), and the computed measurement errors were within the manufacturer's specified tolerances. During the validation process, we collected 30 data points per sample position with the measurement sensor mounted on a rigid mount.. The standard deviation of each measurement value (i.e., X, Y, Z, roll, pitch and yaw) was calculated, and is shown in Table 1. These deviations illustrate that the uncertainties of the laser tracker for measuring the ground truth object are also within the specified tolerances, and justify the utilization of the laser tracker system as a ground truth for evaluating 6DOF pose estimation systems with purported accuracy tolerances of ≥0.15 mm.

4. ALUMINUM MECHANICAL FIXTURE

To compensate for the single-target limitation and setup complexity of the Laser Tracker system, we developed a portable aluminum mechanical fixture ground truth system (*GT2011*) capable of supporting several NIST manufacturing part artifacts simultaneously. These artifacts were designed to represent a quorum of features found in typical manufacturing environments. Each artifact is a modular block with machined features found in real-world manufactured parts. GT2011, shown in Figure 2, was designed to generate repeatable ground truth artifact poses, and then provide this pose data in the form of known homogeneous transformation matrices to researchers for algorithm evaluation. The aluminum construction provides stiff transformations, and limit wear of the fixture over time. A limitation of this fixture is that it requires precision machining capabilities to produce; as such, the cost to produce this ground truth in-house was approximately $4,000.

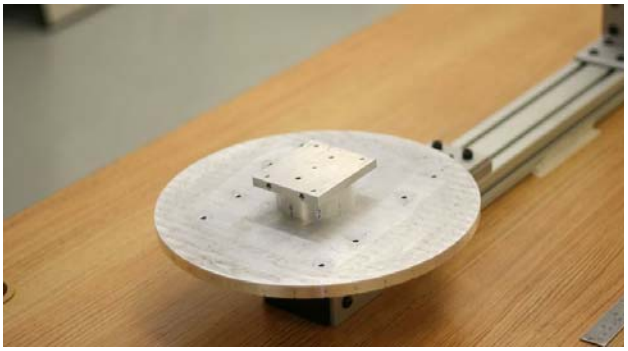

Figure 2. The GT2011 rotation base plate with a mounting plate affixed to the Fixture 0 mechanical offset.

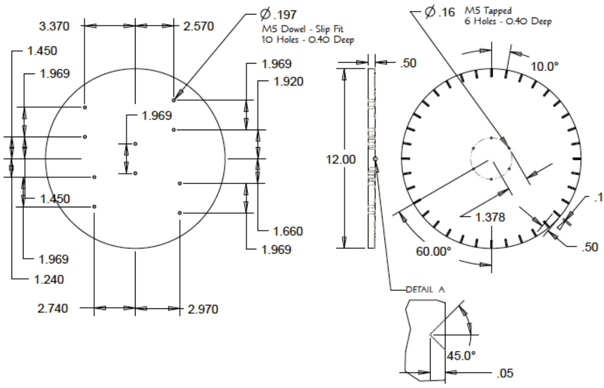

Figure 3. The base plate design of GT2011 featuring five mounting position (left) and 36 rotation presets (right).

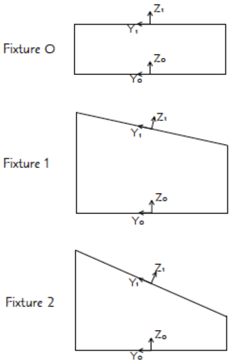

Figure 4. Illustration of the angular tilt offsets generated by the three mechanical offsets.

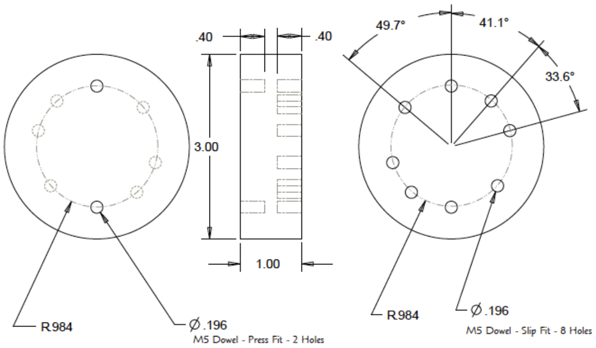

Figure 5. Illustration of the angular rotation offsets generated by each of the three mechanical offset fixtures.

4.1 Design

The design of GT2011 consists of a frame constructed from 8020, a modular engineering system of interlocking aluminum components. The base is used to hold the sensor under evaluation on a vertical arm, with adjustments to vary the sensor horizontal and vertical offsets relative to the rotation plate mounted via a slew bearing to the 80/20 (a modular aluminum framing system) base. The rotation plate (Figure 3), machined at NIST, contains four sets of two alignment holes that accept mechanical offset fixtures (Figure 4) and NIST modular manufacturing part artifacts. Each mechanical offset fixture provides an angular offset as a machined surface for attaching an artifact, and four sets of alignment holes (Figure 5) for attaching to the alignment holes in the rotation plate via dowel pins. The alignment holes enable each offset fixture to be rotated 49.7º, 105.3º, 138.9º and 180.0º. From Figure 4, offset Fixture 0 has a nominal 0º tilt and a vertical (Z axis) offset of 25.4 mm. Offset Fixture 1 has a nominal 12.3º tilt, and a vertical offset of 42.49 mm. And offset Fixture 2 has a 23.8º tilt, and a vertical offset of 34 mm. The plate can also be rotated at 10° increments using a ball plunger quick lock mechanism to produce over 2,300 6DOF positions per artifact.

Up to four artifacts can be placed on the rotation plate at a time for producing artifact occlusions. The fixture's design provides comparatively high accuracy, but has limited range. Relative positioning errors of the ground truth can be attributed to the machining process which is typically accurate to within approximately ±0.02 mm per alignment hole, and inaccuracy associated with the slew bearing tolerances.

4.2 Measured Accuracy

The laser tracker's active target was rigidly affixed to one of the NIST manufacturing part artifacts such that the target was co-centric with the fixture's alignment holes. This artifact was, in turn, mounted on the GT2011 fixture via these integrated alignment holes. For reference, with regard to Figure 3, the center of the base plate is henceforth referred to as TP0, the upper left is noted as TP1, the upper right as TP2, lower left is TP3, and the lower right is TP4. Because TP0 is co-located with the center of the base plate's rotational axis, it is typically utilized at NIST as a reference point for training purposes. It is therefore not evaluated in this study, but instead provides the basis for relative transformation analyses. Additionally, only the position uncertainties of the remaining four TP locations are investigated.

We measured the X, Y and Z coordinates for the laser tracker's active target in each of the four evaluation TP positions (i.e., TP1-TP4) oriented in the zero-rotation configuration. The relative distances between each measurement and the measurement made at TP0 was then computed and compared with the nominal distances based on the original CAD design. In all, 32 data points were taken and averaged at each location to compute the measurement error and uncertainty; the results of these computations are shown in Table 2. Over all four TP locations, the GT2011 fixture

Table 2. Relative Translation Measurement Errors

TP Location	Translation Magnitude Error Mean (mm)	
TP1	0.5479	
TP2	0.3376	
TP3	0.4484	
TP4	0.6281	
	Mean: 0.4905	**Variance:** 0.1257

Table 3. Relative Rotation Measurement Errors

Nominal Angle	Rotation Magnitude Error Mean (degrees)	
33.6º	-0.0099	
55.6º	-0.0083	
49.7º	-0.0179	
	Mean: -0.0120	**Variance:** 0.0051

has an average position uncertainty of 0.4905 mm, with a variance of 0.1257 mm.

Similarly, we took 18 measurements of the laser tracker's active target at each TP position of the laser tracker sensor for half of the eight angular rotation offsets created by the mechanical offset fixtures (because the alignment holes enforce 180º rotational symmetry, only four of the eight nominal rotations need to be evaluated). The relative angle between each adjacent nominal rotation measurement is computed and averaged to compute the measurement error and uncertainty. The results are shown in Table 3. In all, the GT2011 fixture has an average relative Z axis rotation measurement error of -0.0120º, with a variance of 0.0051º. Simultaneous with this evaluation, the tilt errors of the two non-zero fixtures were also measured. The results of these measurements are given in Table 4.

Table 4. Fixture Tilt Measurement Errors (degrees)

Fixture Rotation	Mean Error	Error Variance
12.3º	-0.0862	0.0229
23.8º	0.6387	0.0656

The magnitude of the aforementioned measurement errors has been attributed to mechanical complications from the construction of the aluminum fixture. Because of the strict tolerances insisted upon during the construction of the GT2011 fixture, the fit for the dowel pins is quite tight and can result in extemporaneous angular and vertical position offsets from the nominal value. Care should be taken to insure that the artifacts are seated properly when placed on the base plate to minimize this error. We also found significant play in the slew bearings which will require design modifications to minimize table movement when loaded with artifacts.

5. MDF MECHANICAL FIXTURE

The aluminum mechanical fixture design suffered from a few key limitations, foremost of which was the limit in range and modularity. Specifically, because of its design and construction, the range and values of position offsets was limited to the rotational base, the construction of which constituted the bulk of the cost of manufacturing. In contrast, the MDF mechanical fixture system (*GT2012*) was designed to be an even lower cost ground truth system. The GT2012 design, shown in Figure 3, was driven by the desire to have a broad user base of researchers capable of affording a medium-resolution ground truth system to use for future work in algorithm development and tuning.

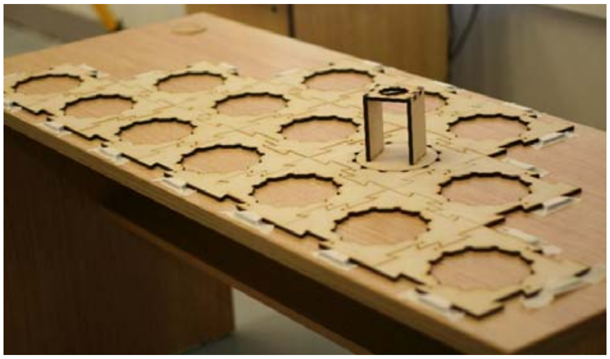

Figure 6. The GT2012 fixture configured for measurement accuracy testing. The base platform can be expaded by attaching additional plates via side interlocks.

5.1 Design

The design criteria used was based on the need for a modular and reconfigurable set of fixturing to support 6DOF positioning of objects similar to the artifact set used with the GT2011 fixture. The GT2012 fixture is designed to be constructed using a light weight, low cost material, and produced using third-party manufacturing services.

The GT2012 system, shown in Figure 6, is constructed from 6.4 mm MDF using a laser cutting process through a web based manufacturing service. It is modular in design such that a base platform is assembled similar to a puzzle, allowing scalability from simple to complex artifact groupings. Each base puzzle piece (Figure 7) accepts a fixture assembly containing two rotational keys, each containing twelve rotational increments, and an angular offset for adjusting Z offset, roll, pitch, and yaw of a mounted artifact (Figure 8). X and Y offsets are adjusted via puzzle piece placement. Additional base pieces are designed for mounting fuducials for calibration of the competitor's measurement systems. The ground truth system made available to researchers for initial testing is comprised of predictable linear and angular offsets. An evaluation ground truth design would be designed using a slightly modified dimensioning scheme using unpredictable offsets.

The cost to produce this ground truth fixture system with the configuration shown in Figure 6 is approximately $400. The fixture's material design and construction does not support the accuracy of the GT2011 fixture, but its modularity compensates for the range limitations of its aluminum counterpart. Relative positioning errors of the ground truth can be attributed to the laser cutting process which produces a kerf of approximately 0.2 mm.

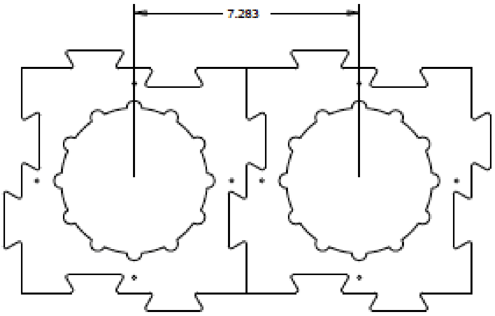

Figure 7. CAD drawing of the GT2012 modular expansion component mounting boards.

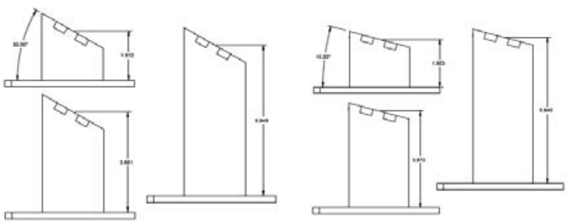

Figure 8. CAD drawing of the GT2012 high-tilt (left) and low-tilt (right) mechanical offsets for three different heights specifications. Not shown is a no-tilt offset option.

Table 5. Relative Translation Measurement Errors (mm) Compared with the Nominal Distance Between Adjacent Mounting Boards

Mean Error	-0.5794
Error Variance	0.3854

Table 6. Relative Rotation Measurement Errors and Uncertainties by Tilt Module (degrees)

Tilt Module	Mean	Variance
No Tilt	-0.0351	0.1147
Low Tilt	-0.0356	0.3203
High Tilt	-0.0498	0.3972
Avg.	0.0402	0.2774

Table 7. Fixture Tilt Measurement Errors (degrees)

Nominal Rotation	Mean Error	Error Variance
15 degrees	0.5901	0.3121
30 degrees	0.1326	0.4176

Table 8. Measurement Accuracy Magnitudes of the Three Evaluated Ground Truth Systems

	Laser Tracker	GT2011	GT2012
Mean Translation Error (mm)	0.015	0.4905	0.5794
Translation Error Variance (mm)	0.0053	0.1257	0.3854
Mean Rotation Error (degrees)	0.03 (active target)	0.1522	0.1686
Rotation Error Variance (degrees)	0.0007	0.0312	0.3124

5.2 Measured Accuracy

To evaluate the measurement accuracy of the GT2012 fixture, we arranged 15 of the modular expansion components described earlier in the configuration shown in Figure 6. The laser tracker was rigidly affixed to the mid-height, no-tilt mechanical offset (shown inserted into one of the expansion boards) and moved to each of the fifteen mounting positions in the zero-Rotation configuration (co-linear with the principle axis of the laboratory table). We measured the X, Y and Z axis coordinates of the laser tracker sensor in each position, and calculated the relative distances between each pose measurement. These distances were then compared against the linear criteria distance of 184.988 mm between the centers of adjacent mounting holes. The results of these comparisons are given in of these calculations are shown in Table 5.

Every mechanical offset fixture integrates two rotation keys—a rotation key base plate for integrating with the modular expansion components, and a smaller key hole to accommodate individual artifact mounting and rotation—each containing twelve rotation increments of 30º, and a preset angular tilt angle. For this study, only the key base plate rotations were assessed for measurement accuracy. To evaluate the rotational accuracy, the relative angular distances between adjacent rotational increments were evaluated and compared with the nominal 30º criteria angle. For each rotational measurement, 30 samples were taken and averaged to calculate the measurement error mean and variance. The results of these calculations are shown in Table 6.

As with GT2011, the mechanical offsets for GT2012 introduce both translational (Z axis) and rotational transformations for a given artifact. For each nominal Z offset (50.0126 mm and 99.9998 mm), three different angular values are introduced: a nominal 0º angular offset ("no tilt"), a nominal 15º offset ("low tilt"), and a 30º offset ("high tilt"). The low and high tilt offsets are illustrated in Figure 8. The six non-zero angular values introduced by the mechanical offsets were measured and compared to the nominal 0º tilt offset. For each measurement, 18 sample data points were taken, and the measurement errors and variances were then calculated. The results of these calculations are show in Table 7,

In contrast with the GT2011 design, the tolerances of GT2012 are far less rigid, and the material properties of MDF allow for faster wear as a function of use and time when compared with the aluminum and steel construction of GT2011. As a result, the measurement uncertainty of the GT2012 fixture increases with use. The low cost of the system, however, permits ready replacement of component parts as they wear.

Table 9. Utility of the Three Ground Truth Systems

	Laser Tracker	GT2011	GT2012
Max number of objects per scene	1	4	Unlimited*
Range (depth)	0 m – 80 m	0.6 m – 2.0 m	Unlimited*
Range (XY)	± 320º azimuth -60º – 77º elevation	0 m – 0.25 m	Unlimited*
Cost (US$)	150 000	4 000	400

* - Theoretical; though, due to the modular design of the fixture, the larger the area spanned by the objects over the fixture, the greater the pose uncertainties.

6. CONCLUSIONS

In this paper we presented three ground truth measurement systems actively utilized at NIST for the evaluation of 6DOF pose estimation systems: a laser-tracker based system; GT2011, a low-cost machined aluminum fixture system; and, most recently, GT2012, a laser-cut, MDF fixture. The laser-tracker ground truth system is used to evaluate the 6DOF pose of a fiducial in Cartesian space, while the two fixture-based systems are intended to provide *a priori* pose data based on known transformations from a reference position via mechanical offsets relative to a given sensor under test. A comparative matrix of measurement errors and variances is given in Table 8.

The evolution of the ground truth systems demonstrate a growing trend in modularity, and an emphasis in lowering cost to make the solutions more accessible to researchers. These are in-line with ongoing standards efforts at NIST, and are being integrated by the ASTM E57.02 standards committee for 6DOF static pose estimation system evaluation. The cost-to-modularity ratios inherent with these efforts are illustrated in Table 9. As was seen, however, a consequence of emphasizing lower cost and modular design is an increase in measurement error and uncertainty.

7. ACKNOWLEDGMENTS

We thank Gerry Cheok from NIST for her technical support in this study. We would also like to thank Tommy Ji of the University of Maryland, College Park for his work in generating conceptual designs of the MDF Mechanical Fixture.

8. REFERENCES

[1] Chang, T., Hong, T., Falco, J., Shneier, M., Shah, M., and Eastman R. 2010. Methodology for evaluating static six-degree-of-freedom (6DoF) perception systems. In *PerMIS '10: Proceedings of the Performance Metrics for Intelligent Systems Workshop*. 2010.

[2] Devore, J. L. *Probability and Statistics for Engineering and the Sciences*. Brooks/Cole Publishing Co., Monterey, CA. 93940. 1987.

[3] Shah, M., Chang, T., and Hong, T. 2009. Mathematical Metrology for Evaluation of 6DOF Visual Servoing System. In *PerMIS '09: Proceedings of the 9th Workshop on Performance Metrics for Intelligent Systems*. 2009.

[4] SMARTTRACK SENSOR. 2007. Automated Precision Inc. http://www.apisensor.com/PDF/SpartTrackeuDE.pdf. 2007.

[5] Nikon Coordinate Measuring Machines. 2011. Nikon Metrology. http://www.nikonmetrology.com/optical_cmm/. 2011.

[6] Vicon Motion Systems. 2011. Oxford Metrics Group PLC. http://www.vicon.com/. 2011.

[7] Madsen, C. B. 1997. A Comparative Study of the Robustness of two Pose Estimation Techniques. *Machine Vision and Applications*. 9, 5-6 (1997), 291-303.

[8] Yoon, Y., DeSouza, G.N., and Kak, A.C. 2003. Real-Time Tracking and Pose Estimation for Industrial Objects Using Geometric Features. In *Proceedings of the IEEE International Conference on Robotics and Automation*. 3473-3478. 2003.

[9] Shakunaga, Takeshi. 1992. An Object Pose Estimation System Using a Single Camera. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1053-1060. 1992.

[10] Wülfing, J., Hertzberg, J., Lingemann, K., Nüchter, A., Wiemann, T., and Stiene, S. 2010. Towards Real Time Robot 6D Localization in a Polygonal Indoor Map Based on 3D ToF Camera Data. In *Proceedings of the 7th IFAC Symposium on Intelligent Autonomous Vehicles*. 2010.

[11] Chang, T., Hong, T., Shneier, M., Holguin, G., Park, J., and Eastman, R.D. 2008. Dynamic 6DOF Metrology for Evaluating a Visual Servoing System. In *PerMIS '08. Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*. 2008.