Summary of NIST's Efforts to Develop Protocols for the Performance Evaluation of 3D Imaging Systems

Geraldine S. Cheok, Alan M. Lytle, Kamel S. Saidi
National Institute of Standards and Technology
100 Bureau Drive, MS 8611
Gaithersburg, MD 20899-8611
cheok@nist.gov, alan.lytle@nist.gov, kamel.saidi@nist.gov

Abstract: This paper presents a synopsis of the effort at the National Institute of Standards and Technology (NIST) to develop standard protocols for the performance evaluation of 3D Imaging Systems. A 3D imaging system is an instrument used to measure rapidly the 3D coordinates of points on an object or within a scene. The effort toward developing standard protocols for performance evaluation began with a workshop at NIST in 2003. Since the 2003 workshop, three other workshops have been held at NIST with the latest one on March 2-3, 2006. The objectives of the NIST workshops were to provide a forum for sharing and discussing efforts in the evaluation of 3D imaging systems and to determine the needs of the 3D imaging community. One outcome of the 2006 workshop was the selection of a standards development organization (SDO) to host the standard protocols for 3D imaging systems.

Keywords: 3D imaging systems, LADAR, laser scanning, performance evaluation, targets, terminology, test protocols.

1. INTRODUCTION

Although the technology for most 3D imaging systems has existed for several decades, the use of these instruments, has only become more established or accepted in the past 10 years, and it is still considered an emerging technology in many industries. In this same time span, the technology for 3D imaging systems has experienced significant advancements. However, there are still no standard test protocols for evaluating the performance of terrestrial 3D imaging systems and assessing the accuracy of their derived output such as 3D models, volumes, and geometric dimensions. This lack of standard test methods is inhibiting a wider market acceptance of these systems, not only in the construction sector, but also in the manufacturing and transportation sectors. Standard test methods for the performance evaluation of 3D imaging systems will provide a basis for fair comparisons of such systems, reduce the confusion regarding terminology, and increase user confidence in these systems.

3D imaging systems are used to rapidly capture (thousands of measurements per second) 3D information of a scene or object. This information is often provided in the form of point clouds with associated color and intensity data. The systems include laser scanners, laser radars, 3D optical scanners, 3D range cameras, and 3D flash LADARs (laser detection and ranging). Sub-classes of these systems of particular interest include those that are ground-based and are capable of capturing a scene that is on the order of a large capital project such as a process plant, construction site, building, or a bridge.

The avenue towards standardization of performance evaluation of 3D imaging systems was a series of workshops. Between 2003 and 2006, the National Institute of Standards and Technology (NIST) held four workshops to address the need for standard methods for the performance evaluation of 3D imaging systems [1, 2]. This paper presents the summaries and outcomes of these workshops.

2. WORKSHOP SERIES

The workshops were attended by manufacturers, users, and researchers from academia and government agencies from the U.S. and abroad. The following sections present summaries of the workshops in chronological order.

2.1 1st NIST Workshop - June 12-13, 2003 [1]

At NIST, the growing use of LADAR[1] technology and LADAR data processing underscores the necessity of an intramural test facility. In keeping with its mission as the Nation's metrology laboratory, NIST is in a position to provide metrology support to both users and manufacturers of LADARs in addition to meeting its own substantial internal calibration needs.

In support of its mission, NIST conducted a workshop directed toward the establishment of a LADAR calibration[2] facility on June 12-13, 2003 in Gaithersburg, MD. The objectives of the workshop were:

– to provide a forum for sharing and discussing the then current efforts in LADAR calibration

[1] The term LADAR was used in the 1st and 2nd workshops but was replaced by the term 3D imaging systems in the 3rd workshop. The general feeling was that the acronym LADAR was unfamiliar to most in the construction/manufacturing communities and was not commonly used outside of the military community.

[2] Based on the 1st workshop, the emphasis was changed from "calibration" to "performance evaluation".

- to determine the types of performance evaluations and test protocols required
- to identify the physical requirements of a calibration facility
- to explore potential plans for the establishment, operation, and location of a LADAR test facility

In general, there was strong support for standardization. An underlying theme in many of the workshop discussions was how to deal with the fundamental question: "Do you really want calibration, or performance assessment/evaluation, or certification?" The terms "calibration", "performance evaluation", and "certification" have similar meanings and have been used, at times, synonymously.

It was felt that calibration is performed to determine the hardware characteristics that enable setting or alignment of instrument parameters to optimal levels.

Performance assessment/evaluation is a voluntary assessment conducted at the request of an end user to determine how well the instrument and the processing software meet the end user's specific requirements. The performance assessment could also include software analysis.

Certification has legal connotations and involves testing of the instrument in accordance with a set of protocols and the results compared against a pass/fail metric. In general, the testing takes place in a certified laboratory and is voluntary.

Because of the large investment involved in acquiring LADAR instruments, users need to have confidence in stated claims or specifications and be confident that what they are purchasing would meet their particular needs. The following measures would aid in building this confidence:

- clarification of manufacturers' specifications to enable meaningful comparisons between various commercially available instruments
- uniform guidelines for manufacturers' specifications, testing, and reporting
- performance testing of individual user-owned instrument upon request at a neutral or independent facility

Although many manufacturers have gone to great lengths to test and evaluate their products, they affirmed the need for quality assurance and uniform specifications such as:

- common set of terminology
- facilitation of "factory floor" calibrations through the use of NIST traceable artifacts and standard procedures
- availability to manufacturers of climate controlled facilities for testing/calibration, particularly, under extreme conditions
- uniformity of specification testing and reporting

For users, an independent facility where one may send an instrument for performance evaluation is desirable. On the other hand, the majority of manufacturers prefer a set of standard protocols and/or artifacts that allow in-house testing rather than a certification procedure. Properties of interest to both users and manufacturers include range, beam pointing, beam size/spread, and processing multiple returns (mixed pixels or phantom points caused by the splitting of the laser beam at edges).

The consensus was that a single facility that would encompass the entire range of LADARs would be impractical. Therefore, a minimum of three kinds of testing facilities was necessary:

- a small, highly climate controlled indoor facility for highly accurate, short range instruments (< 10 m)
- a medium sized, climate controlled indoor facility for instruments with ranges up to 50 m
- an outdoor testing area for long range instruments and for testing in a more realistic environment

While the emphasis at the workshop was on ground-based LADARs, the outdoor facility could be extended for use with airborne LADARs. It was also felt that input from the airborne LADAR community be sought in this "standardization" process, at least during the early stages, as there were similarities between the ground-based and airborne instruments.

In summary, there was almost universal agreement on the need for an independent facility. Three common themes ran throughout and stood out in the discussions. These recurring themes centered on the need for:

- common set of terminology
- standard targets/artifacts/standard reflectivity
- performance assessment/evaluation

2.2 2nd NIST Workshop – March 15-16, 2005 [2]

Based on the findings of the 2003 LADAR workshop, NIST initiated a small, indoor, artifact-based facility for evaluating LADARs. Efforts included procuring a high-resolution scanner, developing potential test artifacts, and acquiring appropriate laboratory space and determining necessary modifications to the space. In determining the types of artifacts, input from LADAR users and manufacturers was essential since the type of artifacts are dependent upon the application. For example, if a geometric model was sought, accurate physical dimensions were important, or if the ability to discern a small object or feature was needed, knowledge of instrument resolution was important. As determined in the 2003 workshop, the latter example points out the need for definitions for commonly used terminology – in this case, what is the meaning of "resolution"?

NIST held a second workshop on March 15-16, 2005 to solicit input from LADAR manufacturers, end-users, and researchers on commonly used terminology and their definitions, measurements of interest, and types of artifacts for use in a performance evaluation facility. The objectives of the workshop were to:

- review and modify preliminary draft definitions of commonly used terms,
- determine the types of measurements and levels of accuracy needed

- determine what artifacts are needed to evaluate the measurement

With regards to terminology, the general feeling among the participants was that it did not matter which definitions [e.g., International Vocabulary of Basic and General Terms in Metrology (VIM), U.S. Guide to the Expression of Uncertainty in Measurement (GUM), ASTM, survey definitions, etc.) were adopted so long as a set of definitions was adopted and accepted by the laser scanning community. A point was made that if standardization was the desired outcome of the current efforts then definitions that are adopted by organizations such as ISO and ANSI should be used.

The terms accuracy, precision, and resolution created some confusion as they have often been used interchangeably. The consensus was that resolution was not encoder resolution but was instrument resolution in terms of depth (range) resolution, horizontal resolution, and vertical resolution. For example, horizontal resolution is the distance two objects have to be laterally separated before one can tell that there are indeed two objects and not one. Some additional terms suggested for inclusion to the list were beam divergence, beam spot size, rated conditions, limiting conditions, registration, and compensation.

The participants expressed mixed feelings about the need to characterize an instrument. Some participants felt that it was necessary to determine the instrument characteristics, such as range uncertainty, while others felt model or project accuracy was more important. However, most felt that knowledge of both instrument characteristics and project accuracy was necessary.

Participants stated knowledge of instrument characteristics as important because when purchasing an instrument, users want a way to fairly compare instruments and have confidence in the manufacturers' specifications. Standard test protocols are required to enable fair comparison between instruments and an educated decision as to the best instrument for a particular application. Additionally, the test protocols used by the different manufacturers to determine their instrument's uncertainty differs, and because of this, the resulting uncertainties would be different; thus, one-to-one comparisons of stated specifications for different instruments are not possible.

Knowledge of the project accuracy was important to users/service providers and facility owners. The basic questions are:

- How accurate is the 3D model?
- How do you verify the accuracy of the model?
- How do 3D models derived from LADAR data compare to those obtained using other methods?

Regarding instrument characteristics, participants ranked range uncertainty as the most important. Other characteristics felt to be important were reproducibility, repeatability, angular uncertainty, and instrument resolution. Other characteristics discussed were data rate and correlation.

In terms of developing test protocols for range uncertainty, the parameters considered to have the most effect on range uncertainty include environmental conditions, target reflectivity, target material, angle of incidence, and distance to target. There was general agreement on the use of planar targets with standard (known) reflectivity. Issues to consider when developing the protocols include:

- The inability of some scanners to acquire a single measurement, i.e., they can only operate in scanning mode.
- The ability of some scanners to average measurements. Averaging would increase the time to acquire a scan and decrease the noise. Therefore, the scan or measurement time needs to be reported.
- The determination of the distance intervals for placing the targets – linear, uniform spacing or random spacing? It was generally felt that smaller intervals be used for targets at ranges between 1 m and 20 m then for targets at ranges above 20 m.
- The inability to de-couple hardware and software.
- The ability to center an instrument over a benchmark.
- How to obtain reference measurements. Currently, LADAR measurements are compared to measurements from traditional survey instruments. The anticipation is that the accuracy of LADAR devices will likely equal the accuracy of traditional survey instruments in the near future.

The definition of angular uncertainty generated some discussion. Was angular uncertainty "how accurately is the angle between two points determined?" or was it "how accurately can the instrument be pointed, i.e., if the angle between measurements was specified to be 0.05º, did the instrument go to 0.05º or did it go to 0.04º?" Participants suggested several protocols for obtaining angular uncertainty for the former definition. However, it was felt that developing protocols for the latter definition would be much more difficult. Angular uncertainty, as defined by the latter definition, allows for a direct means of quantifying the uncertainty of a point in Cartesian coordinates using azimuth, elevation and range measured by a spherical instrument.

Regarding project (or model) accuracy, participants suggested the use of scaled artifacts to allow for indoor and outdoor evaluations. These artifacts could include spheres, tetrahedrons, and cylinders. There was only limited discussion as to what constituted project accuracy. For example, was project accuracy the point-to-point accuracy, the ability to identify an object or features on an object, or the ability to detect misaligned components? Currently, a control network of benchmarks is used to infer or to extrapolate project accuracy.

In the general workshop discussions, there was some discussion on topic of standard data format and metadata. Most of the participants felt that a standard data format would allow for interoperability of the hardware and software.

2.3 NIST/FIATECH Industry Workshop: Assessing the Accuracy of Field Measurement Technologies, July 13, 2005.

A workshop sponsored by NIST and FIATECH was held at NIST on July 13, 2005 to determine industry interest and obtain their input for assessing project accuracy or the accuracy of 3D models (or end product) derived from 3D imaging systems. This need was identified in the 2nd workshop and was reinforced when NIST was approached by FIATECH to collaborate on this activity. FIATECH is a non-profit consortium focused on fast-track development and deployment of technologies for the substantial improvement in the design, engineering, building, and maintenance of capital projects and facilities.

The goals of the NIST/FIATECH project and NIST's efforts are the development of consensus-based procedures for assessing the:

- accuracy of field measurement technologies and equipment
- quality and completeness of the data and meta data
- quality of work processes and procedures
- accuracy of the modeling software

The NIST/FIATECH workshop objectives were to determine the current practices in quantifying resulting model accuracy, scope of the field demos, and project plan.

The group developed an outline of a document that listed the tasks and steps necessary to assess the accuracy of 3D models. The outline of the document follows:

- Guidelines for Readers
- Field Measurement Technology
  - What do customers need?
  - Levels of accuracy required for different types of applications
- LADAR Technology Introduction
  - What is the current state of technology?
  - Why are we looking at it?
  - What is the potential cost savings based upon data quality, time, safety, etc.
- LADAR Quality Assurance
  - What are the sources of error during the data capture and analysis process? (Calibration, field survey control, scanning, registration, modeling)
  - Discussion of Error Budgeting
  - Discussion of Error Sources
    - Calibration
    - Field Surveying
    - Scanning
    - Registration
    - Modeling
  - Methods of Controlling Error (Optimum Work Processes)
    - Calibration
    - Field Surveying
    - Scanning
    - Registration
    - Modeling
- Verification
  - Calibration
  - Field Surveying
  - Scanning
  - Registration
  - Modeling
  - As-built
- Standards and Specifications
  - Concepts for End-Product Performance Specifications
  - LADAR data and meta-data standards
  - ISO 9000 Certification
  - Accreditation / Licensing
- Glossary

The outcome of this workshop was the formation of a FIATECH project[3] - Laser Scanning Measurement Assurance (LSMA). The objective of this project is to evaluate the accuracy between the laser scanner data and the resulting 3D model as well as documenting and sharing current best practices for laser scanner use.

2.4 3rd NIST Workshop – March 2-3, 2006

Based on input from the first two NIST workshops, NIST initiated the development of a small indoor artifact-based test facility for evaluating 3D imaging systems, developed a pre-standard document of terminology for 3D imaging systems, and drafted a pre-standard ranging protocol. The 3rd workshop was conducted to solicit comments and input from instrument manufacturers, end-users, and researchers on the draft terminology and the draft protocol for ranging.

The objectives of the 3rd workshop were to:

- Select a standards development organization (SDO) to house the test methods for 3D imaging systems.
- Finalize the Terminology Pre-standard
- Finalize the Ranging Protocol Pre-standard

The proceedings from this workshop are in preparation, but a brief summary of the workshop outcome is presented.

The workshop participants selected ASTM to house the test protocols for 3D imaging systems. An organizational meeting will be held on June 7, 2006 to formalize the standards committee for this activity.

The Terminology Pre-standard was compiled based on input from the previous workshops. Based on the 2nd workshop, the term LADAR was replaced by the term "3D imaging systems". The definitions for commonly used terms were suggested and any concerns or issues about the suggested definitions were noted at the workshop. The intent was to present this pre-standard along with the noted concerns and issues to the standards committee for their consideration. Terms included in the pre-standard are:

3D imaging systems	angular increment
angular resolution	control points
first return	flash LADAR

[3] http://www.fiatech.org/projects/ijs/lsma.html

frame	frame rate
instrument center	last return
mixed pixels	multiple returns
phantom points	point cloud
point density	registration
registration error	resolution
spatial performance	

The suggested definitions for these terms will be included in the workshop proceedings. The workshop participants also suggested other terms for inclusion to the list:

beam spot size/divergence	compensation
modular transfer function	outliers
panoramic imaging	pixel cross-talk
spatial frequency	

The Ranging Protocol Pre-standard was also presented at the workshop and will be included in the workshop proceedings. The protocol, as proposed, would evaluate instruments to a maximum range of 150 m and would evaluate the absolute distance measurement capability of the instrument. A protocol to evaluate the ranging to longer ranges can be developed at a later time based on the interests of the 3D imaging community. As proposed, the protocol evaluates ranging errors as a function of:

- distance:
  - lesser of (10 to 20) % x ***R*** or 30 m
  - lesser of (20 to 40) % x ***R*** or 60 m
  - lesser of (40 to 60) % x ***R*** or 90 m
  - lesser of (60 to 80) % x ***R*** or 120 m
  - lesser of (80 to 100)% x ***R*** or 150 m

  where ***R*** is the maximum range of the instrument
- target reflectivity (5 levels ranging from 0 % to 100 %)
- angle of incidence (0º, 20º, 40º, and 60º)
- horizontal field-of-view (FOV): 4 user selected levels
  - 0 to ***A***/4
  - ***A***/4 to ***A***/2
  - ***A***/2 to 3/4 ***A***
  - 3/4 ***A*** to ***A***

  where ***A*** is the full extent of the horizontal FOV.

The protocol includes sixty tests. An issue posed to the participants was the number of tests required – the desire is to develop a protocol that was both practical (not too burdensome) but meaningful. Some preliminary feedback from the workshop was that:

- the general concept of the ranging protocol was acceptable
- the protocol should evaluate relative distance and a separate protocol can be developed to evaluate absolute distance
- there is a need to develop separate protocols for measuring a single point on a target and for scanning a target
- the use of planar targets was generally agreed upon

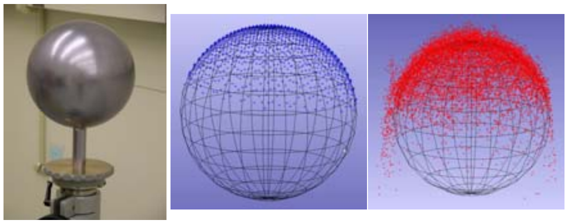
3. NIST INDOOR FACILITIES

Fig. 1 shows the small, indoor, artifact-based facility at NIST. The purpose of this facility is to develop test protocols and metrics for the evaluation of 3D imaging systems. Within this facility, temperature and humidity are continuously monitored and recorded but are not controlled[4].

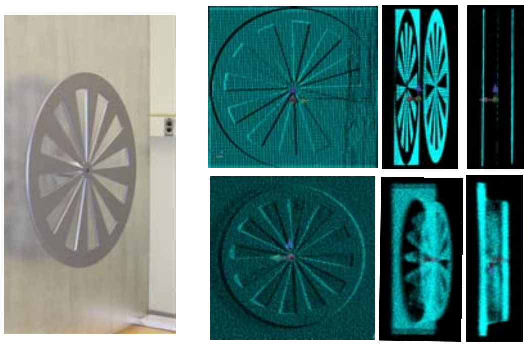
Fig. 1. Indoor, artifact-based facility: 17 m (L) x 5 m (W) x 4 m (H).

Prototype artifacts developed for use in this facility are 152 mm and 203 mm (6 in and 8 in) diameter spheres, a 610 mm (24 in) diameter slotted disc, and a stair artifact. Fig. 2 shows these artifacts and some scans of these artifacts. A 3 m ball bar (Fig. 3) was also manufactured which could be used in the field to determine if an instrument was within specified tolerance. The ball bar consists of a carbon fiber reinforced tube with a 152 mm (6 in) diameter SMR (spherically mounted reflector) at each end. Five 610 mm x 610 mm (24 in x 24 in) square targets with known reflectivity ranging from 2 % to 99 % were purchased for indoor and outdoor use. Another target, 457 mm x 457 mm (18 in x 18 in), consisting of four different reflectivity bands on the same target was also obtained.

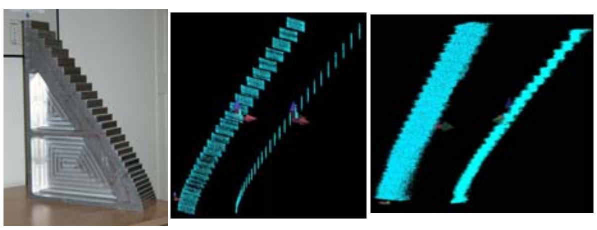
[4] Over a 4-month period, the average temperature was 20.2 ºC with a standard deviation of 0.4 ºC and the average humidity was 23.9 % with a standard deviation of 7.9 %.



a. 203 mm diameter sphere. Images on the right show the point clouds and best-fit spheres.



b. Slotted disc artifact. Images on the right show the front, oblique, and side views of the point clouds.



c. Stair artifact. Images on the right show the point clouds from 2 different instruments.

Fig. 2. Some prototype artifacts and scans from different instruments.

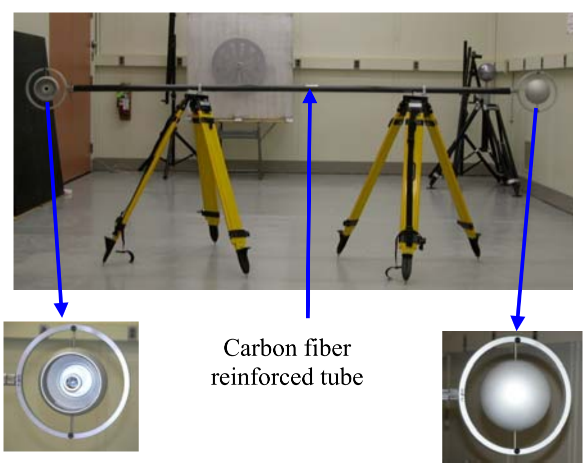


Fig. 3. 3 m ball bar artifact.

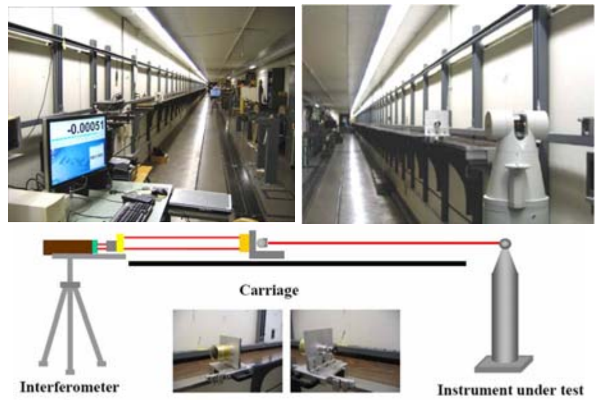


Fig. 4. 60 m Ranging Facility

In addition to the artifact-based facility, an indoor ranging facility[5] at NIST is also available (Fig. 4) that can provide reference measurements with an uncertainty of 10 µm ± 0.5 µm/m (10 µm ± 0.5 ppm). This facility would fall under the medium range facility mentioned in Section 2.1. The facility is temperature and humidity controlled, and the barometric pressure is monitored. A rail system is used to position the targets up to a maximum distance of 61 m. Other artifacts available for use in this facility are 102 mm (4 in) diameter titanium spheres and 102 mm (4 in) diameter SMRs

4. SUMMARY AND FUTURE WORK

NIST efforts, which began in 2003, towards the standardization of protocols for the performance evaluation of 3D imaging systems have included input from manufacturers, users, and researchers. These efforts have resulted in the joint development of a Terminology Pre-standard and a Ranging Protocol Pre-standard and have led to the initiation of the formal process of developing performance evaluation standards for 3D imaging systems through ASTM (beginning in June 2006).

Future efforts include testing and evaluation of the ranging protocol, developing procedures/metrics for the use of the artifacts, determining the uncertainty of the registration process, and planning for an outdoor facility.

REFERENCES

[1] **Cheok, G. S., Ed. [2003]**, “Proceedings of the LADAR Calibration Facility Workshop, June 12-13, 2003,” *NISTIR 7054* (CD-ROM), National Institute of Standards and Technology, Gaithersburg, MD, October.

[2] **Cheok, G. S., Ed. [2005],** “Proceedings of the 2nd NIST LADAR Performance Evaluation Workshop – March 15 - 16, 2005,” NISTIR 7266, National Institute of Standards and Technology, Gaithersburg, MD, October.

---

[5] For more information about this facility, contact Steven Phillips, steven.phillips@nist.gov.