Performance evaluation facility for LADARs[1]

Geraldine S. Cheok[a] and William C. Stone[a]
[a]National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8611

ABSTRACT

The use and scope of LADAR (laser detection and ranging) applications continues to expand as the technology matures. This growth is reflected in the National Institute of Standards and Technology's (NIST) experience with research into the applications of LADARs for construction, manufacturing, and autonomous vehicle navigation. However, standard protocols or procedures for calibrating and testing LADARs have yet to be developed. Currently, selections of LADAR instruments are generally based on the manufacturer's specifications, the availability of standard test procedures would promote more uniform definitions of these specifications and provide a basis for a better informed differentiation between LADAR instruments.

Consequently, NIST's Construction Metrology and Automation Group (CMAG) has conducted exploratory experiments to characterize the performance of a LADAR instrument. The experiences gained in these efforts are summarized in this paper. These experiences also pointed to the need for an internal calibration/evaluation facility at NIST, as well as to the need for the development of uniform specifications and test procedures for characterizing LADARs. As a result, NIST convened a workshop on the establishment of a LADAR calibration facility. Discussions of some issues relating to the performance evaluation of LADARs, facility requirements, and similar efforts are presented in this paper.

Keywords: Artifact, calibration, LADAR, performance evaluation, range measurements, standardization, uncertainties.

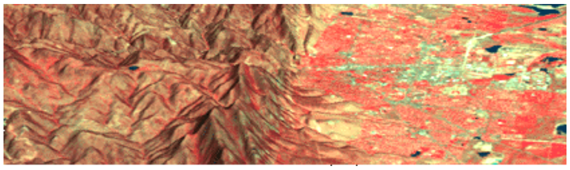
1. INTRODUCTION

Although LADAR technology has been around since the 1960s, the broad use of LADARs developed only within the last decade. This lag was due mainly to prohibitive costs and limited reliability of the early instruments. As the technology matured, however, costs of these instruments have been reduced and reliability has improved.

The applications (Figure 1) for LADARs are widespread and include 3D-modeling, automation, urban planning, mapping, surveying, autonomous vehicle navigation, quality control in manufacturing, global climate monitoring, bathymetry, and homeland security (face recognition, surveillance). The number of applications is seen to increase as technology improves and the size and cost of the LADARs continue to decrease.

At NIST, the growing use of LADAR technology and LADAR data processing underscores the necessity of an intramural calibration and/or performance evaluation facility. These applications include object recognition, tracking terrain changes at a construction site, terrain characterization, collision avoidance, and both indoor and outdoor autonomous navigation. In addition to meeting its own substantial internal calibration needs, providing metrology support to both users and manufacturers of LADARs is in keeping with NIST's mission.

Certain trade names and company products are mentioned in the text or identified in an illustration in order to adequately specify the experimental procedure and equipment used. In no case does such an identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.



Courtesy of Computer Terrain Mapping Inc.

Courtesy of Riegl USA

Figure 1. Examples of LADAR Applications. Terrain mapping and characterization (left). 3D-Models and as-builts (right).

The efforts at NIST aimed at characterizing the performance of a LADAR, establishing a LADAR calibration/ performance evaluation facility, and discussion of the requirements for such a facility are summarized in this paper.

2. INITIAL LADAR EFFORTS AT NIST

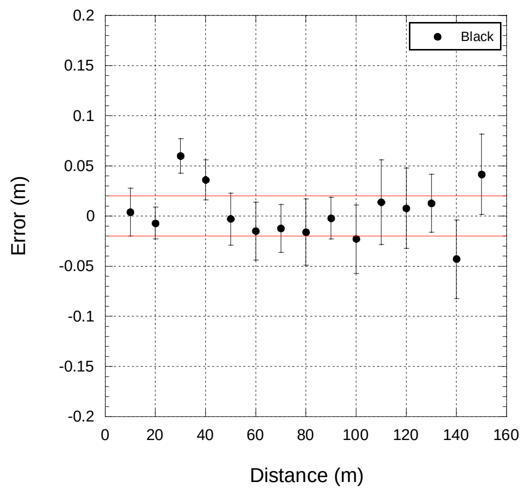
In 1998, the Construction Metrology and Automation Group (CMAG) of NIST acquired a LADAR to determine the viability of its use for construction purposes. The feasibility of using LADARs to monitor construction progress was proven in a live demonstration. The demonstration showed that volume changes to an amorphous object could be readily computed and the changes displayed in a matter of minutes. However, one of the questions raised at that time was "How accurate is the reported volume change?" This question and other questions relating to the accuracy of the reported measurements are the heart of NIST's mission and were, therefore, explored by CMAG in the subsequent years.

Standard procedures for calibrating and evaluating a LADAR are not available. The closest standard is that for a Coordinate Measuring Machine (CMM), ISO standard 10360-2:2001. However, this standard is not applicable for LADARs due to the greater complexities of and different sources of measurement errors for LADARs [1]. The complexities arise because the accuracy of a LADAR is not solely a function of the range measurement and detection hardware, but also a function of the additional hardware necessary for the unit to perform as an imaging system. Also, the different sources of measurement error result from implementing different approaches for range measurement such as time-of-flight, phase-based, and triangulation.

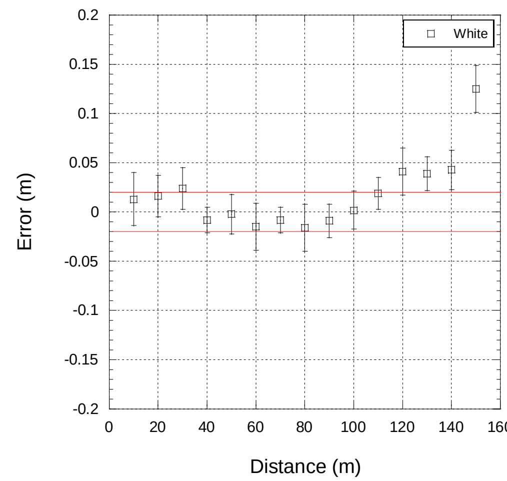
Therefore, to determine the range uncertainty and to evaluate the performance of the LADAR, several exploratory experiments were conducted. These experiments were mainly absolute distance calibrations where the parameters studied were distance (10 m to 150 m – maximum range of instrument), color, and angle of incidence (10º to 90º). No attempts were made to determine the pointing uncertainty of the instrument. This was primarily because of the lack of instrumentation and procedures for the direct determination of pointing errors. Also, based on the manufacturer's specifications, for the distances at which the volume determination was computed, the pointing uncertainty was small compared to the range uncertainty.

The conduct of the exploratory experiments identified several issues. First, as discussed in Ref. 2, the experiments were conducted under less than ideal conditions due to the lack of a dedicated facility for long range calibrations. A shorter range, temperature controlled facility (60 m) was available at NIST but the range was insufficient for the instrument being calibrated and for most instruments that would typically be used at a construction site (≥ 100 m). As there were and are no standard procedures for the test set-up, equipment alignment (that is, initial zero alignment of the LADAR so that the subsequently acquired point cloud can be spatially registered to an external coordinate system), obtaining reference measurements, target size and reflectivity, and the required number of data points, these procedures were developed based solely on best judgment.

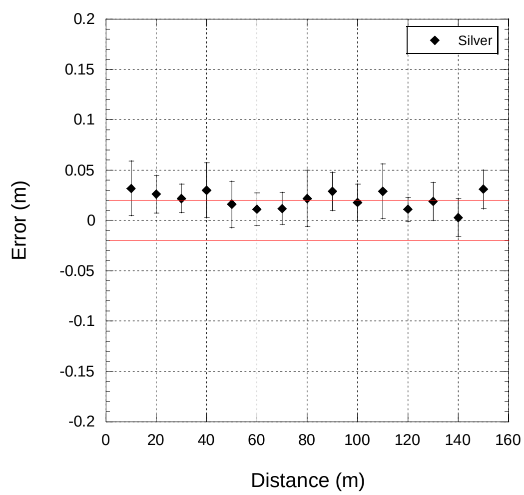
Second, the need for standard targets of known reflectivity became apparent. The targets used were made of thin aluminum sheets that were either painted or unpainted. In some cases the targets were covered with colored paper or reflective sheeting. The targets in the experiments were subjectively classified by color, and with terms such as shiny, smooth, and rough. Assuming that the color, texture, and material of the target all contribute to the reflectivity of the target, a means to objectively quantify the reflectivity of the target is essential. Figure 2 shows some of the results obtained from the range calibrations. In Figure 2, the dashed lines at ± 0.02 m represent the manufacturer's specified accuracy for targets with reflectivity > 80 %. The term "silver" as referred to in Figure 2c corresponds to a target that was unpainted, that is, plain aluminum sheeting. In Figure 2d, an LDP (long distance performance) target was a target that was covered with a prismatic lens sheeting – a highly reflective material that is used for traffic signage.



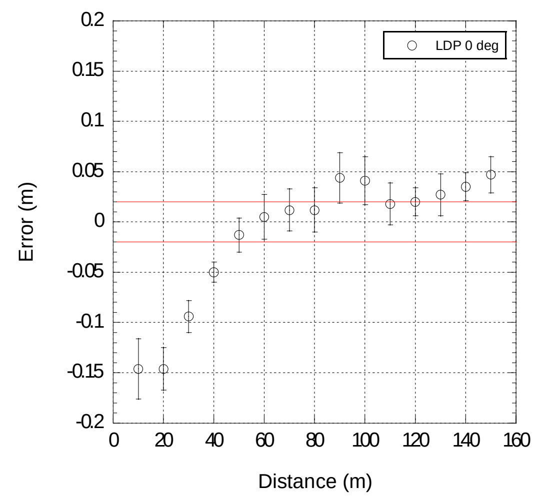
(a) Black Target



(b) White Target



(c) Silver Target



(d) LDP Target

Figure 2. Range Uncertainty (Error bars = 1 σ)

Third, a methodology and test set-up to measure the beam spread needs to be developed as the method described in Ref. 2 was crude at best. The determination of beam spread is an integral part of any method for determining pointing uncertainty. Also, knowledge of the beam spread is also important when determining the resolution of the instrument.

Fourth, two experiments were conducted which involved the scanning of an artifact. These experiments were conducted to test an algorithm to calculate volume and to test an algorithm to register several scans [3]. The artifact used in both experiments was a box made of plywood and painted white. Again, standard artifacts with accurately known dimensions would have been preferable but were not available.

Finally, the correlations between range measurements were investigated. Correlations are needed for assessing error propagation, that is, for estimates of how the uncertainties of primary measurements such as range result in uncertainties of secondary measurements – measurements that are computed from a set of primary measurements. The volume of a scanned artifact is an instance of a secondary measurement.

The experience gained from these exploratory experiments pointed to several issues relating to calibration procedures and facility requirements. Further insights into these issues were solicited from the LADAR community – end users, manufacturers, and researcher – at a workshop reported in the next section.

3. CALIBRATION FACILITY WORSHOP

3.1 WORKSHOP SUMMARY

In view of the experience gained from the exploratory experiments and the expanding use of LADARs at NIST, the establishment of an internal facility to calibrate or evaluate LADARs was a logical next step. Towards this end, the Building and Fire Research Laboratory (BFRL) of NIST conducted a workshop on a national LADAR Calibration Facility on June 12-13, 2003 at the NIST campus in Gaithersburg, MD[4]. The objectives of the workshop were:

- to provide a forum for sharing and discussing current efforts in LADAR calibration
- 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/location of a LADAR test facility

The workshop was attended by a representative cross section of end users and manufacturers as well as private sector and government researchers from Canada and the United States.

Because of the large investment involved in acquiring LADAR instruments, users are in particular need of quality assurance such as:

- 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 instruments upon request at a neutral facility

Manufacturers also expressed support for the objectives of the workshop. Although many LADAR manufacturers have gone to great lengths to test and evaluate their products, they affirmed the need for quality assurance and uniform specifications such as:

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

The LADAR output of main importance to most users is the x, y, z data. As the LADAR output is typically a large point cloud, processing methods should be included in the testing process to provide "end" or "total" performance evaluation. For manufacturers, however, accurate information of the hardware performance is essential for instrument improvement.

The function or purpose of a facility varies due to the different interests of users and manufacturers. For end users, a neutral facility where one may send an instrument for performance evaluation is desirable. On the other hand, the majority of manufacturers at the workshop preferred a set of standard protocols and/or artifacts which allow in-house testing in lieu of a certification procedure. They felt that a certification procedure would involve shipment of each instrument to a neutral facility and this would be very cumbersome and expensive. Properties of interest to both users and manufacturers include range, beam pointing, beam size/spread, and the handling of multiple returns – mixed pixels or phantom points.

The ranges for most commercially available LADARs span 2 meters to several kilometers with uncertainties ranging from a few micrometers to several tens of centimeters. The general consensus was that the establishment of a single facility which encompasses the entire range of LADARs would be impossible. Therefore, three kinds of testing facilities were envisioned:

- 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 an unstructured environment

While the emphasis at the workshop was on ground-based LADAR, the outdoor facility could be extended for use for evaluating airborne LADARs.

Another issue discussed at the workshop was the need for standards. Why have standards? Standards would:

- provide a means for uniform performance evaluation. As the use of or applications for LADAR grows and there are more “naïve” or nascent end users, the ability to fairly compare systems is invaluable. Similarly, when contracting for LADAR services, the ability to insert performance standards into contracts would be very helpful.
- allow end users to conduct their primary business, i.e., manufacture planes, build rail systems, and not have to undertake the task of designing calibration/testing procedures and protocols. Having to devote personnel to this task is costly and often financially difficult for smaller companies.

In general, there was strong support for standardization – one of the participants had commented that a request was made to the then NBS (National Bureau of Standards, now NIST) as far back as the early 1980s for standardization of LADARs. However, it was recognized that standardization involves a long and arduous process. It was pointed out, moreover, that the standardization of a process requires the implementation of proof-of-concept.

In summary, the number of applications for laser scanners was seen to grow rapidly. This being the case, the need for a neutral facility (whether for performance assessment or calibration is yet to be decided) was almost universally agreed upon. There were three common themes that ran throughout and stood out in the discussions:

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

3.2 CALIBRATION, PERFORMANCE EVALUATION, AND CERTIFICATION

One of the fundamental questions that arose in the workshop was the issue of calibration vs. performance evaluation vs. certification. The requirements and specifications of a facility for these three approaches differ. Therefore, an attempt at defining the terms calibration, performance evaluation, and certification was made by the workshop participants.

The terms “calibration”, “performance evaluation”, and “certification” have similar meanings and have been used, at times, synonymously. However, slight differences in the nuances of these terms play a crucial role when establishing a facility for calibration or performance evaluation or certification.

What is calibration? It is generally felt that a *calibration* is performed to determine the hardware characteristics of an instrument to enable setting or alignment of instrument parameters to optimal levels. A more formal definition given by International Vocabulary of Basic and General Terms in Metrology [5] (VIM) is:

> *… a set of operations that establish, under specified conditions, the relationship between values of quantities indicated by a measuring instrument or measuring system, or values represented by a material measure or a reference material, and the corresponding values realized by standards.*
>
> *Notes:*
> 1) *The result of a calibration permits either the assignment of values of measurands to the indications or the determination of corrections with respect to indication.*
> 2) *A calibration may also determine other metrological properties such as the effect of influence quantities.*
> 3) *The result of a calibration may be recorded in a document, sometimes called a calibration certificate or a calibration report.*

Performance assessment/evaluation is a voluntary assessment and would be conducted to determine how well the instrument and the processing software would meet a user’s specific requirements.

Certification has legal connotations and would involve testing of the instrument in accordance with a set of protocols and the results measured against a metric – pass/fail. The testing would, in general, be conducted in a certified laboratory. Product certification is voluntary; however, lack of certification may be negatively interpreted by the marketplace.

The following example for measuring tapes is offered to clarify the difference between certification and performance evaluation [6].

> *An American company wants to sell measuring tapes in Denmark. To do so, the tapes have to meet certain requirements. They meet the requirements and are certified, and the company is given the authority to put the official seal on their tapes. No individual tape needs to be evaluated since they have been certified.*
>
> *In the U.S., the same company simply sells the tapes. The customer either believes the numbers or not. If the accuracy of the tape is important, the customer will request traceability of the measurements. At this point, a higher authority, NIST or a laboratory traceable to NIST, will be asked to calibrate the tape.*

So, one of the first steps towards establishing a test facility is to answer the question: “Do you really want calibration or do you want performance assessment/evaluation or do you want certification?”

In answer to this question, one might ask the following: “What then are the expectations of end users and manufacturers of LADARs?” Based on the definitions above, a performance evaluation facility would be preferred by end users while a calibration facility would be more beneficial for manufacturers.

3.3 WORKSHOP RECOMMENDATIONS

Of the three common issues listed in Section 3.1, the one issue that could and should be addressed immediately is the need for a set of definitions of common terms for LADARs. Therefore, it was suggested that a NIST issued “straw man” of common definitions, addressing in particular, accuracy/ precision/ resolution be sent to the participants for comment.

In addition, the following steps were also suggested:

- contact other professional organizations for possible collaboration/coordination; suggestion was made to include the airborne LADAR community
- conduct a review/inventory/benchmarking of existing facilities
- definition of terms or characteristics of LADAR systems – similarities and/or differences of systems
- list of standard targets and range standards

4.0 CALIBRATION AND PERFORMANCE EVALUATION EFFORTS

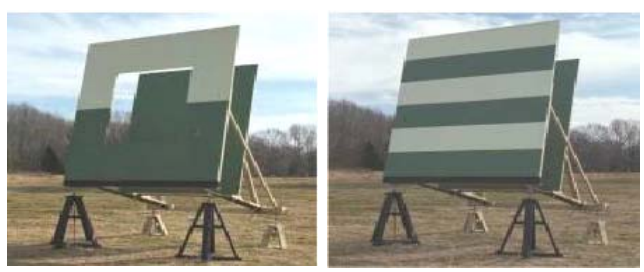
As recommended in the workshop, a review of calibration/performance evaluation efforts was undertaken and is summarized in this section. These efforts include both civilian and military. In these efforts, different test protocols and artifacts were developed based on best judgment of the researchers involved. Some of these facilities, test protocols, and/or artifacts were developed based on the LADAR application.

As stated in Section 3.1, most if not all LADAR manufacturers have developed their own calibration protocols and test facilities, indoors and outdoors, in which to test, calibrate, and evaluate their own products. However, these protocols and/or calibration equipment are often considered proprietary and are not included in the following discussion.

4.1 TERRESTRIAL SCANNERS

Two military facilities that conduct performance evaluation of LADARs are the Robert F. Russell Measurement Facility (Figure 3), U.S. Army AMCOM, Redstone Arsenal and the LADAR Development and Evaluation Research Facility at Eglin Air Force Base. These two facilities primarily evaluate LADARs for potential military applications. The Redstone facility has a 100.3 m (329 ft.) tower which houses a laboratory at 91.4 m (300 ft.). In addition, an elevator can be used as a movable platform for conducting experiments at various elevations from 0 m to 91.4 m (0 ft. to 300 ft.). There are several test areas (ranges varying from 200 m to 1.1 km) surrounding the tower. The researchers at Redstone have developed test procedures, artifacts (2 examples shown in Figure 3b), and software to evaluate a LADAR's potential for object detection and recognition, reconnaissance, foliage penetration, etc. Similarly, the researchers at the Eglin facility have developed test procedures and artifacts to evaluate characteristics of a focal plane array (FPA) LADAR – range accuracy and resolution – in an outdoor facility [7]. They also studied the effects of a moving target, signal strength, and target reflectivity on the range and resolution and the uniformity of the range accuracy and resolution across a scanned image.

a. Tower Housing the Laboratory



b. Targets

Figure 3. Robert F. Russell Measurement Facility.
(Courtesy of J. Grobmyer, U.S. Army Aviation and Missile Command)

The National Research Council of Canada (NRCC) is also involved in the development and calibration/evaluation of LADARs. They have established a temperature controlled metrology laboratory (Figure 4a) for LADAR calibration and performance evaluation [8]. The NRCC facility has also developed calibration procedures/methods and artifacts (Figs. 4b and 4c) for LADAR performance evaluation.

(a) Metrology Laboratory (b) Target Array (c) Artifacts

Figure 4. Metrology Facility at the National Research Council of Canada. (Courtesy of F. Blais, NRCC)

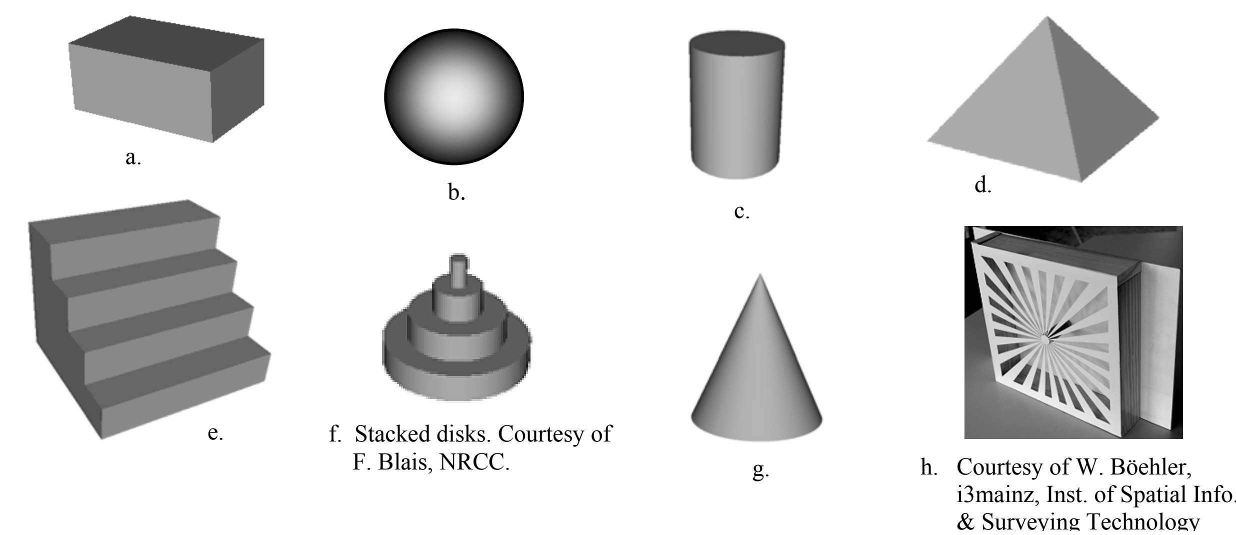
A group of researchers from 3D-MATIC Research Laboratory, University of Glasgow and from the Imaging Faraday Partnership, Sira Limited, initiated a research program in 2001 to establish a facility for assessing the Performance of 3-D imaging systems in the United Kingdom [1, 9]. The objectives of the project are to obtain a better understanding of the factors affecting the accuracy of non-contact 3D imaging systems and to assess a methodology for testing these systems. Two performance characteristics that the group identified as key concerns to the 3D imaging community were *i*.) overall geometric accuracy throughout the measuring volume and *ii*.) sensitivity to surface properties. In the development of a methodology, critical issues that needed to be addressed were ease-of-use and cost-effective to implement.

A performance evaluation of nine commercially available LADARs were conducted by i3mainz, Institute for Spatial Information and Surveying Technology, Germany [10]. Characteristics that were evaluated included angular accuracy, range accuracy, resolution, edge effects, and surface reflectivity. White spherical targets were used to determine relative distance; planar targets were used to characterize range accuracy; spheres in combination with stairs were used to characterize angular accuracy; the target shown in Figure 5h was used to characterize resolution; a pipe was used to evaluate the ability to extract dimensions (the pipe diameter in this case) of an object modeled using the point cloud; and offset plates were used to study edge effects. The evaluations were performed in an indoor environment.

The Gävle GIS Institute of Sweden also has conducted performance evaluations of three commercially available LADARs [11] in an indoor laboratory environment. The evaluation included range accuracy, accurate representation of a 3D object – edge effects and angle of incidence, and the effect of reflectivity on measurement accuracy. A planar, wooden target was used for range accuracy. Three targets were used for evaluating accurate model representation: *i*.) two parallel plates set at a fixed distance apart, with one plate having known geometrical cut-outs, *ii*.) box with different materials on each face, and *iii*.) two metallic plates were used to evaluate the ability to measure and model small angles between surfaces.

Researchers at the University of Ferrara, Italy [12] evaluated a LADAR under field conditions using an historic building as their artifact. The parameters studied were range accuracy as a function of distance and angle of incidence.

Some artifacts that have been or could be used for performance evaluation are shown in Figure 5.



a. b. c. d.

e. f. Stacked disks. Courtesy of F. Blais, NRCC. g. h. Courtesy of W. Böehler, i3mainz, Inst. of Spatial Info. & Surveying Technology

Figure 5. Possible Artifacts for Performance Evaluation.

4.2 NON-TERRESTRIAL SCANNERS

A similar effort towards establishing accepted protocols was initiated by the American Society of Photogrammetry and Remote Sensing (ASPRS) for airborne Lidars[2] (light detection and ranging). ASPRS has formed a committee to develop the "Operational Lidar Guidelines". As indicated by the title, this document is meant to aid service providers and end users by giving guidance for good practice[3]. This effort is much broader in scope than the one being undertaken by NIST and, as indicated by the title, provides operational guidelines. These ASPRS guidelines address six topics[4]:

1. System Specification
   - Description of LIDAR System Components
   - Sensor Data from Manufactures
   - Installation Requirements and Recommended Procedures
   - Sensor Calibration & Validation Procedures
2. Technical Overview and Error Budget
   - Technical Background and References
   - Error Sources
   - Types of Errors
   - Identification and Mitigation
   - Reporting
3. Operations
   - System Calibration
   - In-situ Validation of System Calibration
   - Point Spacing/Density
   - Vegetation Cover Considerations
4. LIDAR Data Processing
5. Safety
6. Metadata.

---

[2] LIDARs are used for measuring range, velocity, and chemical concentrations.
[3] A similar document [13] was developed by the Metric Survey Team of English Heritage for terrestrial scanners.
[4] Only relevant subtopics are listed.

In 2003, ASPRS approved and released the ASPRS Lidar Data Exchange Format Standard. This standard format is intended to make the "exchange, manipulation, analysis and storage of lidar data faster and easier" [14] between hardware manufacturers, software developers, data providers and end users. Draft guidelines entitled "Vertical Accuracy Reporting for Lidar Data" were released in February 2004.

NASA's Stennis Space Center also has initiated an effort to establish an outdoor test range for digital elevation product characterization. The planning and methods for establishing the test range were based on the National Digital Elevation Program (NDEP) guidelines [15] and the North Carolina Floodplain Mapping Program. Five types of digital elevation models (DEMs) or products are defined in the NDEP guidelines: *i*.) Digital surface model (DSM), or first reflective surface – DSM could include treetops, rooftops, or ground surface, *ii*.) Digital terrain model (DTM), or bare earth surface – terrain with vegetation and man-made objects removed, *iii*.) Bathymetric surface (submerged surface of underwater terrain), *iv*.) Mixed surface (some combination of 1st three surface types), and *v*.) Point cloud. The NDEP guidelines are comprehensive (topics include surface modeling, data quality, glossary) and would serve as a good starting point for planning an outdoor facility for terrestrial LADAR evaluation.

5. PERFORMANCE EVALUATION FACILITY

The initial experimental efforts at NIST indicated the need for a calibration and/or performance evaluation facility. Feedback from the workshop discussions supported the establishment for such a facility. The calibration/performance evaluation efforts both in North America and Europe also indicate a need for a neutral facility. The challenges to the establishment of such a facility result from the wide range of LADARs.

Based on CMAG's exploratory experiments and on the workshop discussions, CMAG decided that a performance evaluation facility would be initiated rather than a calibration facility. This decision was also based on the fact that calibration of a LADAR involves knowledge of the internal design of the instrument which is usually not readily available and would involve constructing a facility that is specialized for a particular type of instrument. It is anticipated that range "calibrations" would be conducted at this performance evaluation facility but calibration in the sense that the errors and uncertainties of the range measurements would be ascertained but not the sources contributing to these values.

However, before such a facility can be developed, several steps need to be performed. First, the LADAR characteristics to be assessed must be identified. Then standard test protocols and performance metrics have to be developed. Finally, an architecture of a facility must be determined.

In order to assess the LADAR characteristics, the question that needs to be asked is "What performance is evaluated or sought?" and the answer is "It is dependent on the application."

For example, when assembling manufactured parts, uncertainties relating to relative distance measurements would be of paramount importance. On the other hand, when measuring excavation volumes and determining where to excavate, uncertainties relating to both relative and absolute distances are critical. Uncertainties associated with the sensor (e.g., GPS, INS) used to locate an excavator must also be included when accounting for uncertainties of absolute distance measurements. For volume calculation, the software used to create the surfaces for volume calculation contributes to the volume uncertainty and must be considered. This includes the method of registration (if two or more scans are involved), data point selection, meshing, and data cleaning/filtering. When extracting certain features such as crack widths or irregularities on a surface, knowledge of the scanner resolution would be essential. Resolution in this case is taken to mean the minimum distance between objects and the minimum object depth that is detectable. Two factors influence the resolution – the size of the angular increment and the laser beam/spot size [characterized most effectively by the beam spread function (BSF)]. Edge effects are also influenced by the laser beam size.

In general, several LADAR characteristics of common importance include: range (absolute and relative) uncertainty, resolution, repeatability (closeness of measurements under same conditions), and reproducibility (closeness of measurements under changed conditions). Some factors that influence these characteristics include distance, target reflectivity, and ambient environmental conditions. As some of the current instruments claim micrometer level accuracy, a critical issue to consider is the accuracy of the reference measurements: can a reference measurement that is an order of magnitude better than the instrument being evaluated be obtained?

Once the performance criterion is selected, the procedure to evaluate the performance has to be developed. In developing the evaluation procedures, care has to be taken to ensure that the procedure be inclusive of all scanners to the extent possible. It is envisioned that a set of standard targets and artifacts will be required. The artifacts and targets have to be standardized in terms of spectral reflectivity, size (target size has to be larger than beam size), shape (Figure 5), wavelength (600 nm to 1500 nm), and material (e.g., Invar).

The next step would be the development of metrics to quantify the performance of the LADARs. In some cases, this is a straightforward procedure (e.g., range uncertainty) while more complex in other cases. An example of the latter case is the development of metrics to quantify how well an instrument captured terrain features. The difficulty arises from the fact that the establishment of "ground truth" for terrain is sometimes extremely problematical if not impossible. Another example is the development of metrics to quantify how well a registration algorithm performs. A well-defined artifact may be used to evaluate the algorithm but this procedure may not be applicable for more amorphous objects encountered in, for example, a construction site. Also, the test area for registration evaluation should cover a region that ranges from the minimum to the maximum range of the instrument as registration errors are more apparent at the longer distances.

Initially, the performance evaluation facility would most likely be an indoor artifact facility that could accommodate measurement volumes of about 15 m x 15 m x 10 m (H). This facility would be highly controlled in terms of the environment (temperature, humidity, pressure). Reference measurements would likely be measured using interferometers and laser trackers. A second facility would be required for longer distance calibration (50 m to 100 m) and would likely be a "tunnel" with an automated rail system for positioning the target. This facility may or may not be environmentally controlled. Reference measurements would likely be made using interferometers, laser trackers, and total stations. A third facility could be an outdoor facility that could be used for long range calibrations and performance evaluation in a field environment. Evaluation of sensor performance under actual conditions is an important issue for end users. The outdoor facility could encompass wooded, open, and urban terrains. The effect of the different seasons (amount of foliage cover) could also be studied in an outdoor facility. A set of benchmarks will have to be located throughout the test area for referencing the test instrument and for target placement. The benchmarks would likely be surveyed in and provisions made for rapid generation of high-resolution ground truth immediately prior to a performance evaluation of a test instrument

In addition to the physical facility, a suite of "standard", preferably open source, software is essential for post-processing of the LADAR data and for the generation of an un-biased performance score set. The development of standard data sets will be required to compare different algorithms, determine the limitations of an algorithm, and determine the robustness of an algorithm.

5. SUMMARY

The applications for LADARs have grown in the past decade and continue to grow. Research efforts for the use of LADARs for the construction industry and autonomous navigation have also seen similar growth at NIST – highlighting an internal need for a calibration/performance evaluation facility. However, there are no standard procedures for the calibration or performance evaluation of LADAR instruments.

Exploratory experiments conducted to characterize a particular LADAR, recommendations from a workshop on the establishment of a LADAR performance evaluation facility, and the independent efforts to calibrate/evaluate LADARs support the need for standard procedures for performance evaluations and calibrations of LADARs and a facility in which to conduct these tasks. Challenges to the establishment of a performance evaluation facility include the wide range of LADARs and the development of performance metrics. The performance sought and hence, the performance metric, is highly application dependent.

The establishment of a performance evaluation facility involves the: *i*.) construction of the physical facility, *ii*.) development of definitions of common terms for an unambiguous and rational interpretation of evaluation/calibration results, *iii*.) development of standard targets, artifacts, and test protocols for comparative and repeatable tests, and *iv*.) development of open source software to support the facility.

REFERENCES

1. S. J. Marshall, D. N. Whiteford, and R. Rixon, "Assessing the Performance of 3D Whole Body Imaging Systems," *Proceedings of the 6th Numérisation 3D / Scanning 2001*, Congress, Paris, France, 2001.
2. G. S. Cheok, S. Leigh, and A. Rukhin, "Calibration Experiments of a Laser Scanner," *NISTIR 6922*, National Institute of Standards and Technology, Gaithersburg, MD, September, 2002.
3. G. S. Cheok and C. Witzgall, "Triangular Meshing for Volume Determination," *NISTIR 7078*, National Institute of Standards and Technology, Gaithersburg, MD, January, 2004.
4. G. S. Cheok, (ed.), "Proceedings of the LADAR Calibration Facility Workshop - June 12-13, 2003", *NISTIR 7054*, National Institute of Standards and Technology, Gaithersburg, MD, October, 2003.
5. *International Vocabulary of Basic and General Terms in Metrology*, Second Edition, International Organization for Standardization, 1993.
6. Communication with Mr. Charles J. Fronczek Jr., NIST, Precision Engineering Division.
7. B. Miles, J. Land, A. Hoffman, W. Humbert, B. Smith, A. Howard, J. Cox, M. Foster, D. Onuffer, and S. Thompson, "Field Testing Protocols for Evaluation of 3D Imaging Focal Plane Array Ladar Systems," in *Proceedings of SPIE* Vol. **4723**, Laser Radar Technology and Applications VII, G. Kamerman, Ed., pp. 43-56, 2002.
8. J. –A. Beraldin, S. F. El-Hakim, and F. Blais, "Performance Evaluation of Three Active Vision Systems Built at the National Research Council of Canada," *Proceedings of the Optical 3-D Measurement Techniques III*, Vienna, Oct. 2-4, 1995, pp. 352-361, 1995.
9. S. J. Marshall and J. H. Gilby, "New Opportunities in Non-Contact 3D Measurement," *Proceedings of National Measurement Conference*, Harrogate, UK, 2001.
10. W. Böehler, M. B. Vicent, A. Marbs, " Investigating Laser Scanner Accuracy," presented at the XIXth CIPA Symposium at Antalya, Turkey, 30 Sept. – 4 Oct., 2003.
11. M. Johansson, "Explorations into the Behavior of Three Different High-Resolution Ground-Based Laser Scanners in the Built Environment," *Proc. of the CIPA WG6 Int. Workshop on Scanning for Cultural Heritage Recording*, http://www.isprs.org/commission5/workshop, 2002.
12. M. Balzani, A. Pellegrinelli, N. Perfetti, and F. Uccelli, "A Terrestrial 3D Laser Scanner: Accuracy Tests," *Proc. 18th Int. Symp. CIPA*, pp. 445-453, 2001.
13 J. Mills and D. Barber, "An Addendum to the Metric Survey Specifications for English Heritage – The Collection and Archiving of Point Cloud Data Obtained by Terrestrial Laser Scanning or Other Methods," English Heritage, November, 2003.
14. http://www.asprs.org/news.html
15. NDEP, NDEP Guidelines for Digital Elevation, Draft Version 0.3, National Digital Elevation Program, available at http://www.ndep.gov/TechSubComm.html, 93 pp, April 1, 2003.