

Some Issues Relating to Performance Evaluation of LADARs¹

Geraldine S. Cheok, William C. Stone, Christoph Witzgall
National Institute of Standards and Technology

ABSTRACT

The use and scope of LADAR applications continues to expand as the technology matures. This growth is reflected in NIST's experience with research into the applications of LADARs for construction, manufacturing, and autonomous vehicle navigation. However, standard protocols or procedures for calibrating and testing of LADARs have yet to be developed. Because 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 more rational decision.

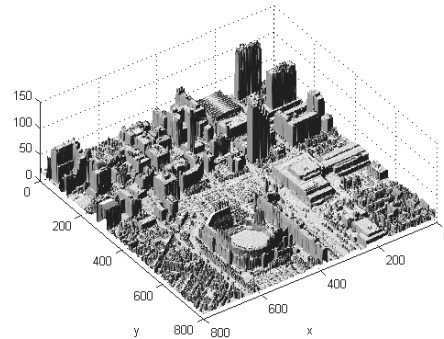
Consequently, NIST's Construction Metrology and Automation Group (CMAG) has conducted exploratory experiments to characterize the performance of a LADAR. The experiences gained in these efforts are summarized in this paper. These experiences also pointed to the need for an internal calibration/testing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

Keywords: *Beam spread, calibration, correlation, LADAR, performance evaluation, range measurements, standardization, uncertainties, workshop.*

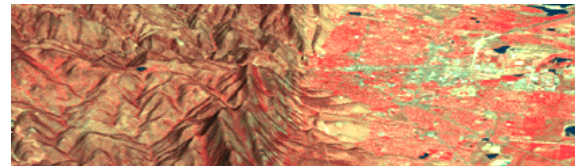
1. INTRODUCTION

Although LADAR (laser distance and ranging) 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.

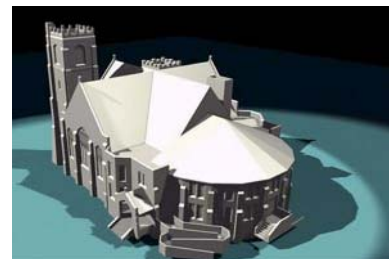
The applications (Figure 1) for LADARs are widespread and include 3D-modeling, commercial automation, urban planning, mapping, surveying, autonomous vehicle navigation, quality control in



Courtesy of J. Turner, Rapid Terrain Visualization ACTD, Ft. Belvoir



Courtesy of Computer Terrain Mapping Inc.



Courtesy of Quantapoint

Figure 1. Examples of LADAR Applications

manufacturing, global climatology monitoring, bathymetry, and homeland security (possibly chemical and biological weapon detection). The number of applications is seen to increase as technology improves and the size and cost of the LADARs continue to decrease.

¹ U.S. Government work and not subject to copyright in U.S.

At the National Institute of Standards and Technology (NIST), the growing use of LADAR technology and 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 such metrology support to both users and manufacturers of LADARs in addition to meeting its own substantial internal calibration needs.

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

2. CALIBRATION, PERFORMANCE EVALUATION, AND CERTIFICATION

What is the intent of a performance evaluation? Is it a calibration, certification, or performance evaluation? 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 VIM [5] 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 interpreted negatively – rightly or wrongly.

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?"

An answer to the question would be, "What then were 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, particularly .

Calibration and determination of uncertainties have been an integral part of NIST's activities, and NIST is also an end user of LADARs. Therefore, to determine the issues involved with LADAR calibrations and performance evaluation, exploratory experiments of a particular LADAR were conducted and are described in the following sections.

3. INITIAL EFFORTS AT NIST

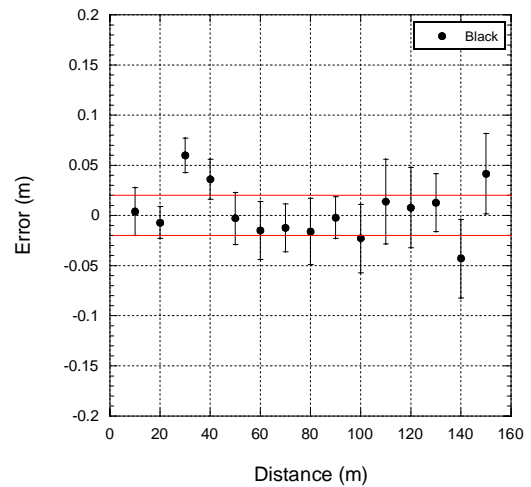
Circa late 1998, the Construction Metrology and Automation Group (CMAG) of NIST acquired a LADAR to determine the viability of its use for construction purposes. The feasibility of using LADARs to monitor construction progress was proven in a live demonstration [4]. The demonstration showed that volume changes to an amorphous object could be readily computed and the changes displayed in almost real time. However, one of the questions raised at that time was “How accurate is the reported volume change?” As this question and other questions relating to the accuracy of the reported measurements are the heart of NIST’s mission, they were investigated by CMAG in the subsequent years.

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 (10° to 90°). 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.

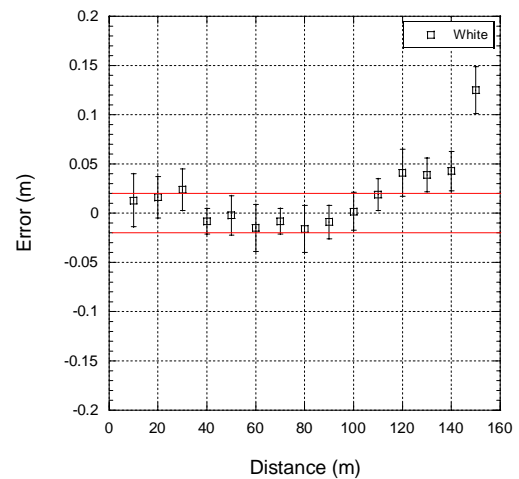
The conduct of the exploratory experiments made several issues clear. First, as discussed in [3], the experiments were conducted under less than ideal conditions due to the lack of a dedicated facility for long range calibrations. A shorter range, temperature controlled facility (60 m) was available at NIST but the range was insufficient for the instrument and most instruments that would be typically used at a construction site. 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 the reference measurements, and the required number of data points, these procedures were developed based on best judgment.

Second, the need for standard targets of known reflectivity became apparent. The targets used were made of flat sheet metal pieces that were 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 Fig. 2, the dashed lines at ± 2 cm represent the manufacturer’s specified accuracy for targets with reflectivity $> 80\%$. The term “silver” as referred to in Fig. 2c corresponds to a target that was unpainted, that is, plain aluminum sheeting. In Fig. 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.

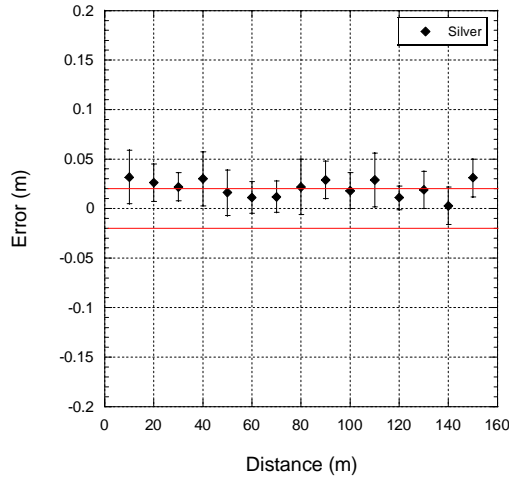


(a) Black Target

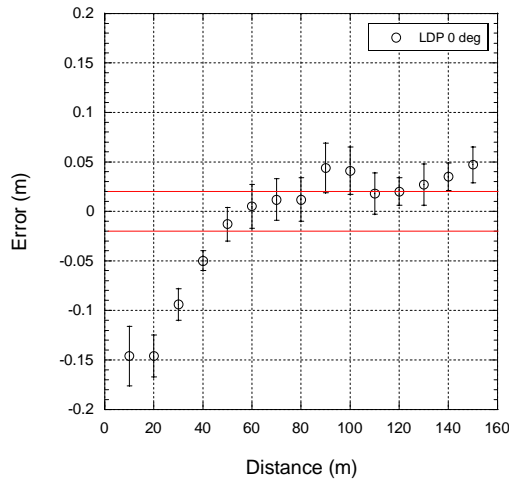


(b) White Target

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



(c) Silver Target



(d) LDP Target

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

Third, a methodology and test set-up to measure the beam spread needs to be developed as the method described in [3] was crude at best. The determination of beam spread is an integral part of any method for determining pointing uncertainty. The need to determine the beam spread or “beam spread function” arose because it was required to deconvolve or to de-blur an intensity image. In an attempt to use the LADAR intensity image to “read” bar codes at various distances, it was found that the images were too blurred to be readable and required image processing techniques to de-blur the image [3]. As mentioned earlier, knowledge of the beam spread is also important when determining the resolution of the instrument.

Fourth, two experiments, not part of the distance calibration experiments but conducted in the same time frame as the distance calibration 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 [2]. 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 estimates of how the uncertainties of primary measurements such as range can be used to determine 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 volume, V , is assumed to be a function of the polar coordinates of the points in the form of:

$$V(r_1, \varphi_1, \theta_1, \dots, r_n, \varphi_n, \theta_n)$$

where

$$\begin{aligned} r &= \text{range} \\ \varphi, \theta &= \text{azimuth and elevation angles} \end{aligned}$$

Assuming that the main contribution to the uncertainty is from the range and not from the pointing angles, a first order approximation of the variance of the volume is given by the standard error propagation formula:

$$\text{var}(V) \cong \left[\left(\frac{\partial V}{\partial r_i} \right)^2 \text{var}(r_i) + \dots + \left(\frac{\partial V}{\partial r_n} \right)^2 \text{var}(r_n) \right]$$

The expression for variance, $\text{var}(r_i)$, as a function of range and angle of incidence has to be calibrated. A simplification would be the use of a constant range variance; this constant value will still have to be calibrated. The derivatives, $\frac{\partial V}{\partial r_i}$ are determined as part of the object modeling process.

The above error propagation formula assumes that correlations are not present. If correlations are known, higher order terms of the series can be taken into account. Therefore, correlations should be determined along with the determination of uncertainties. The incorporation of the error propagation formula into NIST developed algorithms has yet to be implemented.

One of the findings from the experiments was the lack of evidence for temporal autocorrelation (interdependency between consecutive measurements of the same point) for the particular instrument tested. However, random significant spatial correlations between contiguous range measurements were observed which require further investigations to verify. Such verifications might involve additional experiments with alternate experimental designs. Further details of the exploratory experiments and a discussion of the results may be found in Ref. [3].

The experience gained from these exploratory experiments identified 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.

4. CALIBRATION FACILITY WORKSHOP

4.1 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 the National Institute of Standards and Technology (NIST) conducted a workshop on a national LADAR Calibration Facility on June 12-13, 2003 at the NIST campus in Gaithersburg, MD [1]. 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 (approximately 40 participants).

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. However, as the LADAR output is typically a large point cloud, processing methods are to 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 also 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 prefer a set standard protocols and/or artifacts which allow in-house testing in lieu of a certification procedure which would involve shipping each instrument to a neutral facility as 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 5 m 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 airborne LADARs. There was also an opinion that input from the airborne LADAR community should be sought in this “standardization” process – at least these early stages – as there some similarities in the ground-based and airborne instruments.

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 is very helpful.
- allow end users to do 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 – in fact, a specific request was made to the then NBS (National Bureau of Standards, now NIST) as far back as the early 1980s. However, it was recognized that standardization involves a long and arduous process. It was pointed out that the standardization of a process requires the implementation of proof-of-concept – a potential NIST role.

In summary, the applications for laser scanners are seen to be growing 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

4.2 WORKSHOP RECOMMENDATIONS

Of the three common issues listed above was one that could and should be addressed immediately: 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 will be sent to the participants for comment.

In addition, the following steps were also suggested:

- contact other professional organizations for possible collaboration/coordination; may want 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 (a survey of commercially available instruments was published by POB magazine – www.poboline.com)
- list of standard targets and range standards – possibly conduct a survey?

5. FACILITY

The initial experimental efforts at NIST indicated the need for a performance evaluation facility. Feedback from the workshop discussions supported the establishment for such a facility. However, before such a facility can be developed, several steps need to be performed. First, LADAR characteristics for assessment need to 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 application dependent.”

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. The method of registration (if two or more scans are involved), data point selection, meshing, and data cleaning/filtering contribute to the software uncertainty. Finally, 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. An important influence on resolution is the laser beam size [characterized most effectively by the BFS].

In general, several LADAR characteristics of common importance include: range (absolute and relative) uncertainty, resolution, and repeatability. 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 (e.g., flat plate, cube, sphere, tetrahedron), 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 calibration) 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 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 a 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 hardware required for the facility, a suite of “standard”, preferably open source, software will also be required for post-processing of the LADAR data and the generation of an un-biased performance score set.

6. Summary

The applications for LADARs have grown in the past decade and continue to grow – a trend that has been paralleled at NIST. However, there are no standard procedures for the calibration or performance evaluation of these instruments.

Exploratory experiments conducted to characterize a particular LADAR and recommendations from a workshop on the establishment of a LADAR calibration facility highlighted the need for standard procedures for performance evaluations and calibrations of LADARs and a facility in which to conduct these tasks. The performance sought and hence, the performance metric, is highly application dependent.

Prior to the development of protocols or the establishment of a facility, definitions of common terms are needed for an unambiguous and rational interpretation of evaluation/calibration results, standard targets and artifacts have to be developed to enable comparative and repeatable tests, and a review of existing test protocols and facilities need to be conducted.

REFERENCES

[1] Cheok, G. S. (ed.) [2003], "*Proceedings of the LADAR Calibration Facility Workshop - June 12-13, 2003*", in preparation.

[2] Cheok, G. S. and Witzgall, C. [2003], "Triangular Meshing for Volume Determination," NISTIR in preparation.

[3] Cheok, G. S., Leigh, S., and Rukhin, A. [2002], "Calibration Experiments of a Laser Scanner," *NISTIR 6922*, National Institute of Standards and Technology, Gaithersburg, MD, September.

[4] Cheok, G. S., Stone, W. C., Lipman, R. R., and Witzgall, C. [2000], "Ladars for Construction Assessment and Update," *Automation in Construction Journal*, September, pp. 463-477.

[5] International Organization for Standardization [1993], *International Vocabulary of Basic and General Terms in Metrology*.

[6] Communication with Mr. Charles J. Fronczek Jr., NIST, Precision Engineering Division.