Improving the accuracy of surface metrology

Russell Kurtz
Ryder Nesbitt

SPIE

Improving the accuracy of surface metrology

Russell Kurtz
RAN Science and Technology, LLC
609 Deep Valley Drive, Suite 200
Rolling Hills Estates, California 90274
E-mail: russell.kurtz@ranscitech.com

Ryder Nesbitt
Hexagon Metrology, Inc.
7 Orchard Road Suite 102
Lake Forest, California 92630

Abstract. There is a constant search for a more accurate measurement, which generally leads to higher cost, greater complexity, or devices that do not lend themselves to manufacturing environments. We present a method of using statistical sampling to improve metrological accuracy without these undesirable effects. For metrology of flat surfaces and steps between flat surfaces, this method demonstrated precision improvement up to a factor of 55, and accuracy increase of at least a factor of 10. The corresponding precision and accuracy improvements on a spherical surface were both factors of 8. Since this accuracy improvement can be implemented in software, it does not affect the speed of measurement or the complexity of the hardware, and it can be used to improve the accuracy of a wide range of metrology systems. © *2011 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.3602901]

1 Introduction

Metrology can be defined as the science of measurement. This paper specifically discusses dimensional measurement of surfaces, which can be used to determine three-dimensional shape, thickness, and other parameters. This is an important science, from the simple measurements of coins, to more difficult measurements on vehicles, to ultra-precise metrology of semiconductor surfaces. Metrology can be used in automobile manufacturing,[1] aircraft assembly,[2] high-resolution lithography,[3] and nondestructive evaluation.[4] In general, however, there is a tradeoff among various capabilities of a metrology system, particularly among the parameters of system complexity, measurement accuracy, and measurement speed.

It is possible to improve the accuracy through direct trade-offs. For example, increasing the time for a single measurement can reduce the noise in the measurement; covering a larger area with a single measurement can also reduce the noise. These methods, however, fail in specific cases; under conditions of high vibration, increasing the measurement time can actually increase the noise, and if a certain resolution is needed, increasing the measurement area beyond this reduces the measurement accuracy.

The method described here is a version of statistical smoothing. On a flat surface, for example, if N measurements of height are taken at different locations in the plane of the surface, those measurements can be averaged to increase the height precision by a factor of $\sqrt{N}$. On a smoothly curved surface, similar precision improvements can be achieved—indeed, as long as the nominal shape of the surface is known, increases in precision are easy. Increases in accuracy, however, are not quite so simple; only if the random errors induced by the metrology system are independent and zero-mean will accuracy increase as much as precision. Even when the errors have a bias, significant increases in accuracy can be obtained; in these cases, however, precision is improved more than accuracy.

2 Technology Tradeoff

Almost any surface metrology system can have its accuracy improved through statistical methods. Most such systems produce a number of points describing the x (horizontal—left and right), y (distance—front and back), and z (height—up and down) measurement of the surface (a point cloud). Statistical methods can reduce noise by combining a number of measurements at a point—and these measurements can be separated in time or in space. In many cases, such as an environment with a significant level of vibration, averaging a number of points taken at the same time produces a better result. If, for example, the vibration has a small amplitude of 125 μm at a frequency of 20 Hz, a measurement sample time, or time between measurements, of 1 ms results in an inherent measurement inaccuracy of ± 5 μm due to the surface motion; longer periods are even worse. If the time between measurements is 1 ms and there are 100 measurements, this small vibration will show up as a measurement error of up to 250 μm. Since the vibration may not be random, time-based averaging techniques may not remove this error.

The first step in improving the accuracy of the metrology system is to select the technology with the best combined attributes of accuracy and speed. For measurements in an environment with significant vibration, a comparison of technologies (Table 1) shows white light stereovision with a projected pattern (WLSPP) to be the best choice; if vibration is not a factor, the highest resolution comes from interferometry. All the results presented in this paper use WLSPP to generate the point cloud, although the statistical smoothing can be applied to any measurement technology that creates a point cloud.

The values in Table 1 are taken from experimentation using typical devices, and all accuracies listed are at 3σ (99.7% of all measurements fall within $\pm 3\sigma$). Devices that use physical contact, such as the coordinate measurement machine (CMM), can increase the depth of field by increasing the size of the machine. Tomography can also be scaled to larger sizes by using larger machinery, but optical alignment becomes difficult.

Table 1 Potential metrological technologies.

Technology	Sample rate	Accuracy	Depth of field
Interferometry	1 point/1 ms	$\pm$ 5 nm	500 nm
Tomography	1 point/1.3 s	$\pm$ 2.1 μm	10 cm
CMM	0.1 mm^2/5 s	$\pm$ 1.4 μm	30 cm
CMM on portable arm	3 mm^2/2 s	$\pm$ 24 μm	2 m
Pulsed laser radar	1 point/5 ns	$\pm$ 5 mm	> 10 m
Laser scanner	1000 mm^2/5 s	$\pm$ 18 μm	10 cm
Sonar	250 mm^2/10 μs	$\pm$ 350 μm	2 m
WLSPP	2500 cm^2/1 ms	$\pm$ 18 μm	23 cm

As can be seen from Table 1, the compared technology with the best accuracy is interferometry. Our example of interferometry is a simple Mach–Zender system using laser illumination at 532.1 nm; other interferometric systems, such as shearography, speckle interferometry, and holographic interferometry, have other limitations on speed and/or accuracy. For full accuracy, Mach–Zender interferometry is a point (or small area) measurement system, making it potentially too slow to use for large areas or in a high-vibration environment. Interferometry, as well, has such a small depth of field that a large step cannot be measured; it is possible for vibration to move the target enough that the interferometric measurement is no longer valid. Tomography—broadband interferometry with low coherence, so that a signal only appears when the interferometer arms are very nearly the same length—falls into the same category; although it can measure a large depth of field in a single measurement, it is very slow (due to the need to scan a mirror, changing the optical length of one arm), and each measurement is a point rather than an area. The CMM has excellent accuracy and tiling capabilities, but is immobile and slow (our comparison machine was the Brown and Sharpe Global Advantage); a portable version of a CMM, using optical encoders to determine position (for example, a Romer arm), is nearly a factor of 20 less accurate.

The next step in accuracy—at $\pm$ 18 μm in a single measurement—is shared by the laser scanner (such as the Leica T-Scan) and WLSPP (we tested the Hexagon WLS400 and Optigo 200). A moiré interferometer would also fit in this category, while a structured white light scanner (like the Capture3D ATOS) uses time-dependent phase-shifting to accomplish the same results described in this paper with space-dependent technology (see Sec. 3.1). The T-Scan laser scanner is portable and its position is tracked with excellent accuracy by a separate laser tracker. The required motion to cover the scanned area slows this method to where it can be affected by vibration. The other tested technologies are not high resolution, and many are slow. The WLSPP, then, emerges as the base technology for this measurement, with high accuracy within a single shot, reasonable tiling accuracy, the largest area for a single sample, and sample time short enough to significantly reduce the effects of vibration. It should be noted, however, that the statistical methods described in this paper can be applied to any metrological

system that produces an array of points in space (a point cloud), as long as the error sources (both of the measurement and from external causes, such as vibration) are known.

3 Theory and Method

3.1 *Statistical Sampling*

As mentioned in Sec. 1, there is an inherent tradeoff among complexity, speed, and accuracy. By using statistical smoothing, it is possible to increase accuracy through an increase of sample time (reduction of sample rate). It is possible as well to increase measurement accuracy by increasing system complexity. Examples of this include improving the shading measurement capabilities in interferometry, employing a higher ultrasonic frequency in sonar, and increasing the number of pixels per camera in WLSPP. In addition, using structured white light stereovision with multiple samples, in particular using a phase shift mask,[5] can produce accuracies as great as $\pm$ 2 μm over the field of view, at the expense of taking several seconds to make a single measurement.

The statistical technique described here is slightly different. A WLSPP system defines an *x-y-z* coordinate system, with points defined on, for example, a 250-μm square pattern (defining *x* and *y*), with accuracy of $\pm$ 18 μm in the *z* direction. In many cases, there is no need for samples spaced that closely in *x* and *y*; the actual spacing required depends on the curvature of the surface and the measurement accuracy needs. We found that a typical commercial curved surface only required samples every 5 mm to allow excellent surface fitting; this is a factor of 20 greater than the spacing produced by the metrology system directly.

The statistical smoothing technique requires at least an approximate knowledge of the surface shape, whether provided by an *a priori* estimate of this shape or through knowledge of the surface design (nominal). Measurements are then taken to determine the as-measured (actual) surface. The deviation of the actual from the nominal is minimized by adjusting the measured location of the surface, resulting in an increase in both precision and accuracy for the actual value of the surface at a point in the center of the area being sampled.

To determine the actual improvement, each measurement is defined to contain three components:

$$z(x, y) = z_0(x, y) + z_d + z_r, \tag{1}$$

where z_0 is the actual value, z_d is a deterministic error (constant at that point), and z_r is a random error. Further, it is assumed that each value of z_r is completely independent of each other value; in other words, it is a truly random error, it is noise. A nominal value at each point, $z_n(x, y)$, is defined, whether from a known or estimated surface shape or from the known design. The difference between the measured value and the nominal value is

$$\Delta z(x, y) \equiv z(x, y) - z_n(x, y)$$
$$= z_0(x, y) - z_n(x, y) + z_d + z_r. \tag{2}$$

The actual value of interest is the difference between the actual value and the nominal value,

$$\delta z(x, y) \equiv z_0(x, y) - z_n(x, y). \tag{3}$$

If $N = 4ab$ measurements surrounding the point (x_0, y_0) are taken, in (for example) a square pattern, and the difference

values are summed, the result is

$$\sum_{n=-b}^{b} \sum_{m=-a}^{a} \Delta z\,(x_m, y_n)$$

$$= \sum_{n=-b}^{b} \sum_{m=-a}^{a} [\delta z\,(x_m, y_n) + z_d + z_r]$$

$$= \sum_{n=-b}^{b} \sum_{m=-a}^{a} \delta z\,(x_m, y_n) + N z_d + \sum_{i=1}^{N} z_r, \quad (4)$$

where the three parts of the measurement have been separated. The first, double sum is a collection of N measurements of δz. The inside sum runs from $-a$ to a, a total of $2a$ steps; the outside runs from $-b$ to b, for $2b$ steps. The complete double sum, then, includes $4ab$ measurements, defined above as N.

Since z_d is constant (or at least nearly constant in the measurement area), summing it over N steps results in $N z_d$. Since z_r is random, its location does not matter, and the $4ab$ measurements of the double sum can be replaced by a single sum with N samples.

If the double sum in Eq. (4) is divided by N, the result is the average value of δz over the area in question. The second term, when divided by N, is just the constant error—which can be removed, for the most part, through calibration (except when it is an actual surface deviation from nominal). The single sum, divided by N, is the expected error of the measurement,

$$\overline{z_r} = \mu_r + \frac{\sqrt{\sigma_r}}{N}, \quad (5)$$

where μ_r is the mean expected error and σ_r is its standard deviation. Typically (and ideally) μ_r will be close to zero, so the smoothed value becomes

$$\frac{1}{N} \sum_{n=-b}^{b} \sum_{m=-a}^{a} \Delta z\,(x_m, y_n) \approx \overline{\delta z_{ab}} + \frac{\sqrt{\sigma_r}}{N}. \quad (6)$$

So, for example, if an area of 40×40 samples ($N = 1600$) is taken, with an anticipated normal distribution of the random error, with the $\pm 18 \mu$m individual measurement error (at 3σ) of a commercial WLSPP, the expected error of the sampled measurement is $0.15\,\mu$m (6.0 μm standard deviation divided by $\sqrt{1600} = 40$ samples). We assume a normal probability distribution because the Central Limit Theorem indicates that most large random error distributions approximate a normal distribution.[6] The anticipated 3σ error range, then, has been reduced from $\pm 18\,\mu$m to $\pm 0.45\,\mu$m. Note that this does not require a specific surface shape, just that the nominal shape is known. In addition, if the general form of the surface shape is known, it is possible to use statistical curve-fitting techniques to form a nominal surface. Finally, by applying the statistical smoothing just to the differences between nominal and as-measured surfaces, the surface shape becomes unimportant to the noise reduction method.

3.2 Implementation

The statistical smoothing method, then, reduces random noise by the square root of the number of samples averaged. The samples to be averaged must eliminate deterministic errors, to the greatest extent possible. Once this is accomplished, the remaining noise can be reduced. The smoothing method is straightforward, and easy to implement in software. The following steps can be used for a measurement system that uses (x, y, z) triplets, where the location is given by x and y and the surface measurement is z.

1. Collect a point cloud of measurements—the actual values.
2. Determine the expected surface function—the nominal values. (This may be done by mathematical fitting, using a known surface, calculating the CAD surface value at each point in the measured point cloud, etc.)
3. Generate a residual dataset of points $(x, y, \Delta z)$, with Δz equal to the actual minus nominal value for each point.
4. Determine the area over which to smooth the data. (This area may be, for example, a circle, square, or any shape that is easy to implement for the specific task.)
5. Generate a subset of the "residual" dataset that includes all the data points needed, including the smoothing area. [For example, if the smoothing area is a circle whose radius is 1.5 mm, the subset will be those points in the residual dataset such that the (x, y) locations are at least 1.5 mm from the edge of the point cloud.]
6. For each (x, y) point in the subset, calculate the mean value of Δz that is within the smoothing area centered on this point. That is the Δz_s value—the smoothed surface value—in the new point definition $(x, y, \Delta z_s)$. The set of points $(x, y, z_n + \Delta z_s)$ is then the new smoothed point cloud.

3.3 Limitations

The reduction of noise in the data does have a cost. When points are averaged in a measurement area, adjacent points are no longer statistically independent. For spatial sampling, N points are averaged to get a reduction factor of $\sqrt{N}$ in the noise inherent in the z measurement; to a first approximation, the effective resolution is reduced by this same factor in both x and y. For example, the initial point cloud measurement of the sphere (Section 4.2) had points in a grid with 0.221 mm spacing; in effect, after smoothing, the (x, y) resolution was 1.77 mm. This is not exact; the statistical effect of one point on another given distance away is approximately the number of points that fall within the measurement areas of both points, divided by the number of points in each area. Thus, for example, in the same sphere example as above, since each area of measurement was a circle of radius 1 mm, points separated by > 2 mm have no relationship to each other at all. Points separated by 1 mm are 39.1% correlated; adjacent points are 86.0% correlated. Points separated by the calculated smoothed resolution are 14.5% correlated.

4 Experiment

4.1 Measurements of a Plane Surface

We used a WLSPP metrology system (CogniTens Optigo 200) to measure a known flat surface, specifically a gray test flat, manufactured by QS Grimm, GmbH. This block was in certification, defined to have surface flatness $\pm 6.2000\,\mu$m over the center 10.8×1.8-in. area.[7] We tested a strip near the center, 7-in. long and 1.5-in. wide. A scan of this strip is shown in Fig. 1. The scan contains ~ 9000 points. The scale is in micrometer deviation from predicted flatness.

Over the range scanned, where the nominal errors are $\pm 6.2000\,\mu$m, the measured 3σ deviations from flatness are
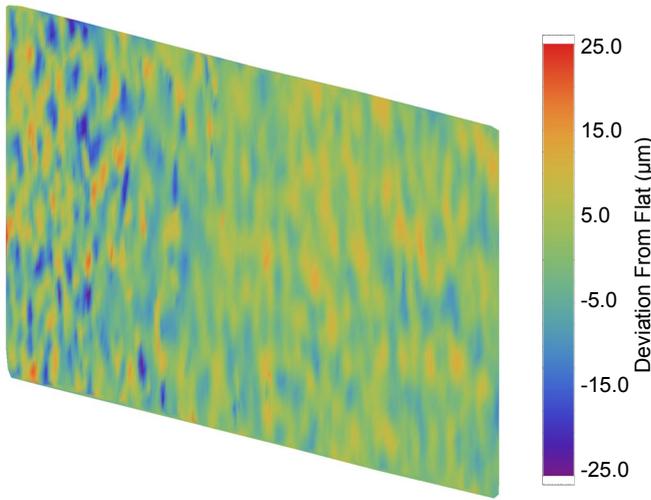
Fig. 1 Metrology of a flatness artifact.

Fig. 3 Comparison of surface deviation before (a) and after (b) statistical smoothing.

± 20.0 µm. This accords with the stated metrology 3σ error of ± 18 µm in one shot, combined with the nominal deviation. The measurement scan was then divided into 14 rectangular sections (seven horizontally, two vertically) for statistical analysis. As described in Sec. 0, the measurement area is a rectangle 1 in. × 0.75 in., each containing ~ 640 points. In this case, each measurement area was resolved to a single point.

The results are shown in Fig. 2. The mean values of the deviations from flatness (in green) average about –0.3 µm (short dashed line), while the standard deviations (in blue) average 4.8 µm (long dashed line). From these measurements we can determine statistically that the 3σ error of the WLSPP was closer to ± 14 µm than the stated ± 18 µm, and the deviation of the surface from flat was about 2.7 µm over each measurement area.

4.2 Measurement of a Sphere

Next, the accuracy improvement technology was used to study a sphere whose diameter was specified to be 1.5000 ± 0.0001 in., but whose surface was stated smooth only to ± 0.001 in. The WLSPP metrology system (in this case, a Hexagon WLS400) was used to scan a section of the upper half of the sphere. To define the nominal surface, a best-fit routine was executed; the radius was known to
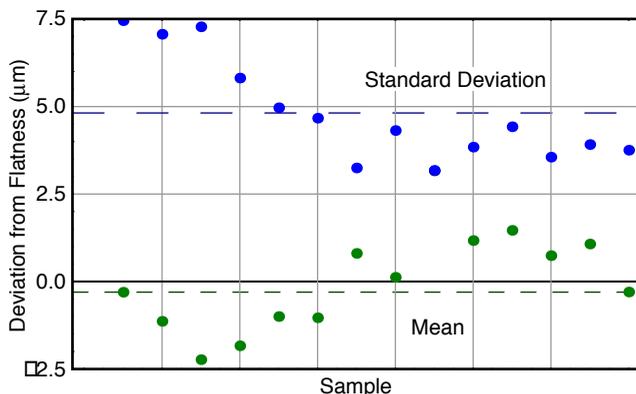
be 19.050 mm (0.7500 in.), reducing the complexity of the fitting routine. The sampled area was bounded by a circle whose diameter was 27.20 mm, and the measurement area was a circle whose radius was 1 mm; the smoothed area, then, was bounded by a circle whose diameter was 25.20 mm. Based on the point cloud density, there were on average 64 measurement points in each measurement area.

The statistical process definitely smoothed the residuals (Fig. 3). The residuals, defined by z [actual minus nominal at each (x, y) point], are plotted with the scale in micrometer; the (x, y) positions are in millimeter. The residual standard deviation—the rms deviation of the actual surface from the nominal—was 18.4 µm in the raw data, reduced to 15.8 µm in the smoothed data. Statistically, this indicates that the 3σ measurement error was significantly better than claimed by Hexagon—approximately ± 9 µm, half the stated value of ± 18 µm—and that the standard deviation of the surface (compared to nominal) was about 15 µm.

Based on the estimated standard deviation of 15 µm, the surface should (statistically) have 9.6% of its points outside the stated smoothness of ± 25.4 µm (± 0.001 in.). The points that fall outside this range are shown in Fig. 4, with the measured deviation, Δz, in µm and the position in mm. Points that are high (> 25.4 µm above the nominal) are red; points that are low (> 25.4 µm below the nominal) are blue. The raw point cloud contained 15.5% of points outside the stated range; smoothing reduced this to 10.6%. (Assuming this 10.6% is the correct number, the calculated standard deviation of the surface is 15.71 µm, close to the estimated value of 15 µm.)

4.3 Step Gauge Test

A block with a simple half-inch (12.7-mm) step was fabricated to determine the accuracy improvement of step measurements (Fig. 5). This was made out of low-expansion stainless steel.

To ensure accurate testing, four areas on the block (lightened areas marked U1, U2, L1, and L2) were measured on
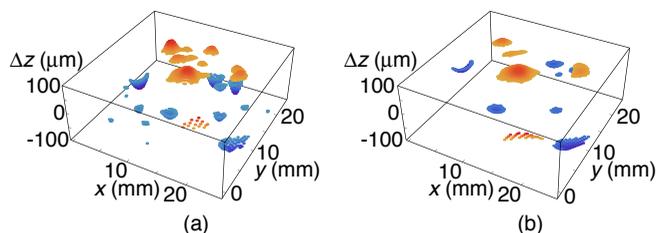

Fig. 2 Statistically-smoothed flatness measurements.

Fig. 4 Areas outside the specified surface smoothness before (a) and after (b) smoothing.

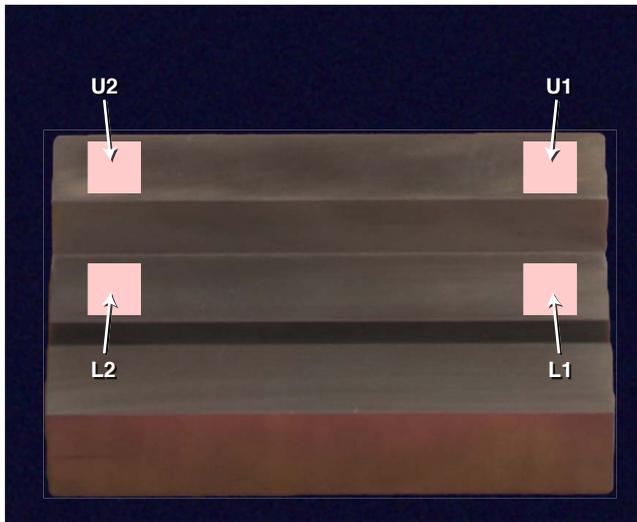**Fig. 5** Step gauge block.

a CMM system with 3σ measurement accuracy ± 1.7 μm. These measurements showed that the surfaces were not quite flat, and that the actual steps from U2 to L2 and from U1 to L1 were not exactly the 12.700 mm expected. It was discovered that the lower level was angled slightly with respect to the upper level, with ~ 50 μm variation from the highest to the lowest points in the measurement areas. The upper level showed surface rippling of ± 7.5 μm (Fig. 6).

The step from U1 to L1, as measured by the CMM, was 12,769.8 ± 2.1 μm (listed error includes both measurement and reporting errors). Raw data from the WLSPP (Optigo 200) produced a step of 12,772 ± 18 μm. Statistically fitting planes to each patch as measured by the WLSPP (since the surfaces are supposed to be flat), then calculating the step from these planes, the step is calculated to be 12,771.78 ± 0.34 μm. This value is just within the step as calculated from the CMM measurements, plus the CMM error range. The second step, from U2 to L2, was measured by the CMM to be 12,679.1 ± 2.1 μm. The WLSPP raw measurement was 12,680 ± 18 μm. With statistical sampling this came to 12,680.22 ± 0.31 μm, again within the uncertainty range of the CMM. Since the accuracy was improved from an uncertainty of ± 18 μm (limited by the WLSPP) to ± 1.7 μm (limited by the CMM), we claim an accuracy improvement of at least 10×. From a repeatability standpoint, the WLSPP exceeded the CMM, having 3σ precision 6.5× better. The gauge study demonstrated that the statistical sampling used (with 256 points each in the upper and lower measurement
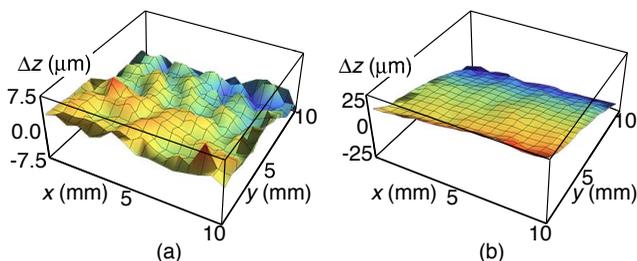
patches) reduced the estimated measurement error to 1.8% of the single-measurement error.

5 Summary and Conclusions

We have demonstrated a significant improvement in surface measurement accuracy by applying statistical methods to metrological point clouds. This can be considered a data smoothing technique, in that it reduces the random errors inherent in any measurement. Using WLSPP metrology, with two different metrology systems, this method was able to improve measurement accuracy of a test flat by a factor of five—to below the flatness specification; of a step measurement by a factor of 55—to better precision than a CMM and with comparable accuracy; and of a sphere by a factor of eight—in both precision and accuracy.

The specific optimization method described in this paper was based on the anticipated requirements for measuring large areas, including steps exceeding 10 mm, in a high-vibration environment. These needs led to the choice of WLSPP for the metrology system, and to the statistical sampling regions (from 64 to 640 samples). In essence, using the large number of samples reduced the resolution in *x* and *y* in order to enhance the accuracy in *z*. Doing so reduced the expected 3σ error of the metrology system from ± 18 μm to ~ ± 2 μm, with precision as accurate as ± 0.31 μm.

References

1. M.-A. Beeck and W. Hentschel, "Laser metrology – a diagnostic tool in automotive development processes," *Opt. Lasers Eng.* **34**(2), 101–120 (2000).
2. K. Vakil, "Automated outer mold line (OML) Measurements And Control," in *Proccedings of the Coordinate Metrology Society Conference* (2010).
3. V. A. Ukraintsev, M. C. Tsai, T. Lii, and R. A. Jackson, "Transition from precise to accurate critical dimension metrology," in *Metrology, Inspection, and Process Control for Proc. SPIE* **6518**, 65181H (2007).
4. R. M. Kurtz, "Directed Acoustic Shearography," in *Photonics in the Transportation Industry: Proc. SPIE* **7675**, 76750B (2010).
5. R. Paschotta, "Phase shift method for distance measurements," http://www.rp-photonics.com/phase_shift_method_for_distance_measurements.html (2010).
6. M. G. Bulmer, *Principles of Statistics*, Dover Publications, Inc., New York, New York (1979).
7. J. Moossa, "Test Report, Inventory No. EBN280× 50× 30/142," QS Grimm GmbH, Karlsruhe, Germany, p. 2 (2007).

Russell Kurtz is the chief scientist at RAN Science and Technology, LLC. He helped advance the field of shearographic NDE, developed the DAS method of locating buried defects, and led the team that applied laser-acoustic vibrometry to detection of buried IEDs. His recent achievements include development of a high-efficiency solar concentrator, fabrication of an automated NDE system, and advancement of interferometric vibrometry. His research interests include applications of lasers, nondestructive evaluation, and the application of automated vibrometry to machine vision. His recent work has concentrated on nanomotion measurement, MWIR solid-state lasers, and aircraft NDE. He invented or co-invented 17 technologies, including directed acoustic shearography and enhanced surface metrology. He is a life member of SPIE and AOC, a senior Member of IEEE, and a member of OSA, DEPS, AAAS, and ASNT, has contributed to more than 20 conference proceedings, and has published over a dozen papers in refereed journals.

Fig. 6 Flatness of the upper (a) and lower (b) level areas of the step gage block.

Ryder Nesbitt is a senior applications engineer at Hexagon Metrology. He received his BS in Physical Sciences from Cal Poly and his MS in spatial imaging from MIT. He worked as a research and application scientist in holographic optical storage, developed electro-optical metrology systems, and applied 3D optical digitization to photogrammetry. He has been project manager on several SBIR and IR&D projects, and has published papers on holographic and spatial imaging, including high-resolution techniques. At Hexagon, he has worked in integration, support, and engineering of the newest high-resolution optical metrology devices, including CogniTens and V-Stars. He has developed unique applications using these devices in photogrammetry, 3D metrology, and maintaining a digital thread of information throughout a range of measurements, including defining software interfaces and user interfaces.