Accuracy assessment and interpretation for optical tracking systems

Andrew D. Wiles[*a,b], David G. Thompson[a,b] and Donald D. Frantz[a]

[a]Northern Digital Inc., 103 Randall Drive, Waterloo, ON, Canada;
[b]University of Waterloo, 200 University Avenue, Waterloo, ON, Canada

ABSTRACT

Highly accurate spatial measurement systems are among the enabling technologies that have made image-guided surgery possible in modern operating theaters. Assessing the accuracies of such systems is subject to much ambiguity, though. The underlying mathematical models that convert raw sensor data into position and orientation measurements of sufficient accuracy complicate matters by providing measurements having non-uniform error distributions throughout their measurement volumes. Users are typically unaware of these issues, as they are usually presented with only a few specifications based on some "representative" statistics that were themselves derived using various data reduction methods. As a result, much of the important underlying information is lost. Further, manufacturers of spatial measurement systems often choose protocols and statistical measures that emphasize the strengths of their systems and diminish their limitations. Such protocols often do not reflect the end users' intended applications very well. Users and integrators thus need to understand many aspects of spatial metrology in choosing spatial measurement systems that are appropriate for their intended applications. We examine the issues by discussing some of the protocols and their statistical measures typically used by manufacturers. The statistical measures for a given protocol can be affected by many factors, including the volume size, region of interest, and the amount and type of data collected. We also discuss how different system configurations can affect the accuracy. Single-marker and rigid body calibration results are presented, along with a discussion of some of the various factors that affect their accuracy. Although the findings presented here were obtained using the NDI Polaris optical tracking systems, many are applicable to spatial measurement systems in general.

Keywords: Spatial measurement, accuracy, calibration, rigid bodies, active markers, passive markers

1. INTRODUCTION

The mid 1990's saw a revolution in the operating theater as spatial measurement systems and medical images were combined to create image-guided surgery (IGS). Many early researchers used the Optotrak measurement system, manufactured by Northern Digital Inc. (NDI) of Waterloo, Ontario, Canada. In 1996 NDI launched a smaller and lower cost system, which has become the system of choice in the IGS market. Significant improvements have been made to the Polaris over the years, including the merging of active-only and passive-only tracking systems into the hybrid tracking system, the development of better light emitting diode markers (LEDs), the improvement of optical distortion compensation models, the creation of larger tracking volumes, and the addition of much new functionality to the Application Programmers' Interface (API). However, the cornerstone requirement for any measurement system remains whether its accuracy is sufficient for the user's intended applications.

For the Polaris position sensor, customer feedback has indicated a widely held misconception that tracking tools with passive markers is inherently less accurate than using tools equipped with active markers. This may arise, in part, from the inability to clean passive spheres in a repeatable manner, which limits them to single

[*]Email: awiles@ndigital.com, Telephone: 1 519 884 5142, Website: www.ndigital.com

usage, and so discourages exhaustive testing with various appropriate configurations. But a major contribution stems from users often not fully understanding some of the key aspects of the representative statistics used to qualify the system's performance. Many users have at best a rudimentary knowledge of some statistical principles based on ideal Gaussian distributions, but the specifications for spatial measurement systems are usually based on the distance error magnitudes, which are inherently non-Gaussian, and more importantly, are seldom distributed spatially in a uniform manner.

We examine some of the key aspects concerning characterization and calibration of optical measurement systems, using the Polaris as our example. Section 2 briefly examines one procedure used to characterize Polaris position sensors and discusses volumetric calibrations obtained from the characterization data themselves. This is followed in Section 3 by a discussion of the interpretation of the representative statistics obtained from the characterizations. Section 4 shows some calibration results obtained for single markers and rigid bodies, and compares the results obtained with passive markers to those obtained with active markers for a typical Polaris position sensor. Section 5 contains a short discussion on application accuracy.

2. CHARACTERIZATION METHODOLOGY

Spatial measurement systems typically use complicated mathematical models to convert their raw sensor data into corresponding 3D positions. These models incorporate various parameters that describe the systems' underlying physical attributes, such as the lens focal length, the lens distortions, and the sensor transformations for optical systems. Although some of these parameters could be determined directly from physical measurements or engineering data, doing so does not generally provide the highest levels of accuracy required for position sensors used in IGS. Thus, such systems are usually characterized to determine their model parameters. Optical tracking systems can be characterized by moving markers throughout their measurement volume in a representative manner according to some convenient reference, whose accuracy is sufficiently better than that of the systems being characterized. For example, a coordinate measuring machine (CMM) can be used to move the markers to a number of accurately known positions. The reference positions and their corresponding sensor data can then be used to determine the model parameters, typically with some appropriate fitting algorithm that minimizes the error between the transformed reference data and the sensor data. The quality of the characterization procedure can be estimated from various measures of the model's quality of fit.

While the assessment of characterization quality is important to manufacturers and is one of the criteria they use to ensure that their systems meet their required quality measures, users are more interested in the systems' performance with regard to spatial accuracy. These measures can be obtained from calibrations, which assess the quality of the systems' reported 3D positions, or rigid body 6D poses. Since the sensor data used to characterize a system can be converted to corresponding 3D and 6D data afterward, once the model parameters have been determined, the converted data can be easily compared to the reference data to provide one type of calibration. Since the calibration is being performed on the same data set that was used to determine the model parameters themselves, such calibrations must be used with caution. For example, certain systematic errors due to environment factors can be compensated by a delicate balancing of the model parameters, which can result in misleadingly low characterization errors, but much higher calibration errors when the system is used in the field, where the corrupting factors are not present. In general, better calibrations are obtained from independent data sets.[1] We refer to calibration errors obtained directly from characterization data as 'characterization errors' to emphasize the close connection.

The characterization procedure for NDI's electromagnetic tracking system, the Aurora, has been previously described,[2] and the procedure for the Polaris position sensor is similar.[3] A CMM is used to move a single LED accurately in a grid of reference positions throughout the Polaris's measurement volume. Several samples are taken at each grid point and averaged to reduce the noise. To ensure that enough data are collected to determine the model parameters with sufficient accuracy, over 900 grid points are used for the Polaris's standard volume, and over 1500 points for its pyramid (or open) volume. The quality of the marker used for characterization is very important, since it has to be representative of the markers that users will typically use. Any specific systematic errors inherent in the characterization marker can corrupt the model parameters as they are determined during the fitting procedure, leading to performance degradations when the system is subsequently used with other markers. This close coupling of the markers with the position sensor complicates the calibrations, since the evaluations assess them both together, whereas separate assessments are usually desired.

For calibration data obtained from grids of several points, such as those obtained from characterization data, the spatial errors at each measured point can be determined by aligning the grids and comparing the measured positions $\vec{r}_m$ to their corresponding reference positions $\vec{r}_r$ on a point-by-point basis as $\vec{\epsilon}_i = \vec{r}_{r_i} - \vec{r}_{m_i}$. Various statistical measures can be calculated for this set of errors as part of the assessment. For example, the commonly cited overall volume root-mean-square (RMS) distance error can be determined for the N points by

$$\epsilon_{\text{RMS}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{\epsilon}_i \cdot \vec{\epsilon}_i)} \tag{1}$$

Similar relations hold for the x, y, and z components. Other important statistical measures of the error distribution include the bias (mean error or median), its spread (standard deviation), and confidence intervals (CI) such as the 95% interval. The reporting of maximum errors is frowned upon, unless the protocol explicitly defines a fixed number of data points N, since the maximum error is not a robust statistic — it tends to increase as N increases, since those errors occurring very infrequently are more likely to be encountered with larger data sets. The 95% or 99% CIs are better indicators of the larger errors in the distribution tails.

3. INTERPRETATION OF CHARACTERIZATION RESULTS

The accuracy specification for the Polaris states that at the time of characterization, the overall volume RMS distance error obtained by stepping a single marker throughout the volume is less than 0.35 mm. Much care must be exercised when attempting to relate this very specific statistical measure to a more general application accuracy. First, the Polaris tracks tools comprised of several markers, while this measure is based on a single marker. Also, application accuracy involves many other considerations, such as rigid body design, rigid body characterization, rigid body tracking algorithms (wired or wireless), dynamic motion, the use of markers different than the ones used to characterize the system, and the distance between the rigid body probes and reference tools. Despite these limitations, single-marker characterization results do provide a common measure for all Polaris position cameras that is independent of rigid body considerations, and so they correlate better with other protocols that use arbitrary rigid bodies than would be the case if a specific rigid body had been used instead.

Another limitation of the overall volume RMS distance error and other representative statistical measures that manufacturers typically present in their marketing material is that much of the underlying information that is necessary to properly assess a given system is lost or hidden. In an ideal case, where the position component errors are free of systematic bias, follow a normal distribution, and are spatially distributed uniformly throughout the measurement volume, the overall volume RMS distance error is a good indicator of the typical error magnitude. But most tracking systems do not meet these requirements, since they typically have substantial systematic errors that do not fit a normal distribution well and are not spatially distributed in a uniform manner. Figure 1 illustrates this for a Polaris position sensor by showing the distance errors that were obtained from its characterization data in four formats of varying detail. Plot i) of Figure 1 shows the spatial dependence of the distance errors (note that even this representation has missing information — the distance error at each grid point is itself a 1D reduction of its underlying 3D error vector, namely its magnitude). The plot clearly shows that errors are mostly uniform within a given xy-plane, except at the upper right corners, and generally increase with the distance from the camera ($-z$). This type of information can be very useful for certain applications. For example, users measuring the length of a predominately 1D object such as a long bar would obtain substantially better results with the object oriented in an xy-plane than they would with the object oriented along the z-axis. Plotting the distance errors instead as a 1D plot as a function of the sequence in which they were collected (plot ii) results in the loss of much of the spatial information, but still shows the general z-dependence, and from the plot's periodicity we can infer that the larger errors are at the volume edges.* In plot iii) of Figure 1, the distance

*The point data were collected starting at the back of the volume, progressing through the same xy-plane, and then moving forward to the next plane; the peaks on the sequence plot correspond to transitions from one plane to the next.

error distribution is plotted as a frequency histogram. The distribution is clearly not normal, as it is skewed heavily to higher errors. This type of distribution is expected because the data being examined are distance errors, which by definition are positive.† Finally, plot iv) of Figure 1 shows some of the representative statistics that describe much of the error distribution, but even this minimal description is further compromised, since most manufacturers typically only quote one or two of these statistics.

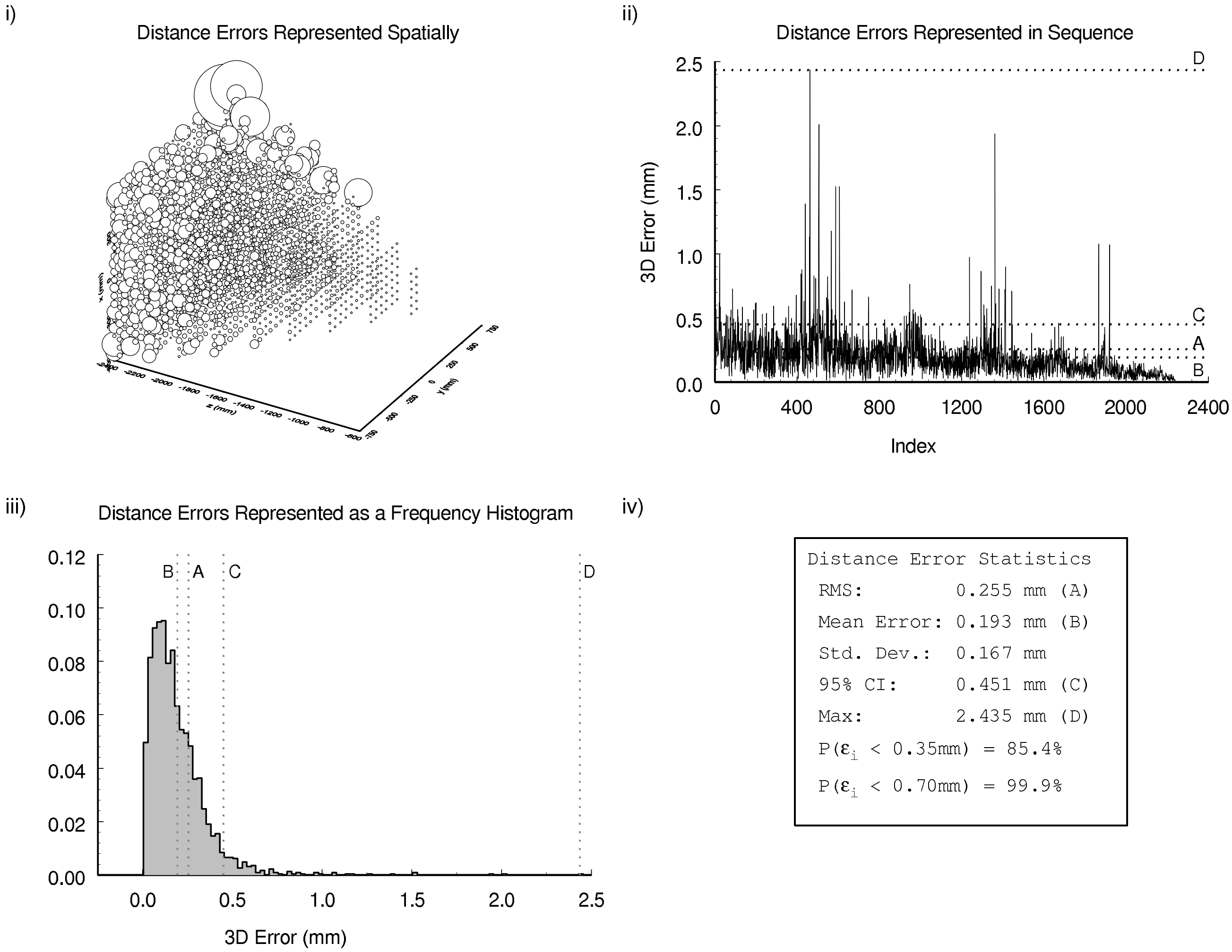

Figure 1. Typical distance error profiles for the Polaris position sensor at the time of characterization. The sequence of plots, from the full spatial representation in i) to the final statistical summary in iv), demonstrates how increasing simplicity and clarity come at the expense of continued loss of information. Plot i) represents the distance errors spatially at each reference position (x_r, y_r, z_r), with the error magnitudes proportional to the corresponding circle diameters. In ii), the distance errors are plotted as a function of the sequence in which they were collected (from the back of the volume to the front). While the 3D spatial information has been lost, the general dependence of the error on the z-depth is still evident, as is the increased error at the edges. In iii), all of the spatial information has been lost, but the skewed non-Gaussian nature of the underlying error distribution is clearly evident (the distribution is related to the general Maxwell probability distribution that governs distance errors[4]). The distance errors are then reduced to the statistical summary in iv), from which only one or two values (typically the overall volume RMS distance error) are presented to the user. These results are also tabulated in Table 2 and compared to other calibration results.

†The distribution is similar to a Maxwell distribution,[4] but the tail in this case is substantially more pronounced because of the dependence of the errors on the z-distance.

When presented with a marketing brochure specification, such as the overall volume RMS distance error for the Polaris being 0.35 mm, many users treat the value as a type of quasi-maximum, naively assuming that most of their application errors would have smaller magnitudes, some slightly higher, and just a few perhaps two or three times greater (the rule of thumb that about 95% of the errors in a normal distribution lie within ± 2 standard deviations is often mistakenly extended to RMS values and non-normal distributions by many users). The first ambiguity arises from the specification being a general one that applies to all of the systems being manufactured. Conservative manufacturers treat such a specification as a threshold, guaranteeing that each of their systems sold has a lower RMS value at the time of manufacture, so that most systems typically have substantially lower values. Other manufacturers treat the specification as a "typical" value, with some systems being better, but others being worse. But even if users are given the actual RMS values for their systems, they can still make unwarranted assumptions. The dotted lines labelled A in Figure 1 represent the RMS value for this Polaris position sensor, and as can be seen in plots ii) and iii), while most of the errors lie within twice the actual RMS value obtained for this system, there is a very small number of outliers having values of several multiples of the RMS, with the maximum error (line D in Figure 1) being an order of magnitude greater than the RMS. Users often want to know a system's maximum distance error, in addition to the RMS value, anticipating that such a specification would set the threshold for the largest error they would encounter in their applications. Figure 1 also shows why such a generic "maximum error" cannot be stated unambiguously. The maximum error in this case is not representative of the error distribution (it is barely visible in the histogram in plot iii), and had the number of grid points been halved, or the grid spacing altered so that those large-error points on the edge had not been included, the maximum error would have been considerably smaller (on the other hand, doubling the number of grid points would have likely found an even larger error). Confidence intervals such as the 95% or 99% CIs are robust statistics, since they vary little with N once the number of points becomes sufficiently large, and so are the preferred indicators of the larger errors that users can expect.

As an alternative to confidence intervals, some users prefer to know the probability that their measurements will fall below the specified accuracy (or some other appropriate threshold). If the error distribution is similar to a known theoretical distribution, then the probability can be estimated from the distribution's characteristic parameters, such as its mean and standard deviation. For example, assuming there are no gross systematic errors in the 3D position error distributions (the means $\mu_x = \mu_y = \mu_z = 0$) and that the errors are uniformly distributed spatially (the standard deviations σ_x, σ_y, and σ_z are constant throughout the volume), then the Maxwell probability distribution can be used to estimate the probability that an error will fall below a given threshold,

$$P\left(\epsilon_r < R_{\text{spec}}\right) = \int\int\int_{V_s} f\left(x, y, z\right)\, dx\, dy\, dz \tag{2}$$

where,

$$f\left(x, y, z\right) = \frac{1}{\sqrt{\left(2\pi\right)^3 \sigma_x \sigma_y \sigma_z}} \cdot e^{-\left(x^2/\sigma_x{}^2 + y^2/\sigma_y{}^2 + z^2/\sigma_z{}^2\right)/2}. \tag{3}$$

For the Polaris, this would only provide a rough estimate, since the standard deviations have a significant z-dependence; alternatively, the value $P\left(\epsilon_r < R_{\text{spec}}\right)$ could be obtained directly from the measured data, provided a sufficiently large number of grid points had been measured. The threshold probability estimates for the specified Polaris accuracy (0.35 mm) and for double that value are shown in plot iv) of Figure 1. They were calculated according to Eqs. 2 and 3 using the overall volume values obtained for σ_x, σ_y, and σ_z. For this system, more than 85% of the distance errors are smaller than the specified threshold of 0.35 mm, and only about 0.1% of the errors are larger than twice the threshold. This result is typical of NDI ceramic markers, whose good characteristic behaviour results in very little marker-to-marker variation. This implies that users can use such characterization results to infer single-marker performance for arbitrary markers of the same type. This is not the case for many other marker types, which have much more marker-to-marker variation, thus limiting the usefulness of extending their characterization results to generic single-marker performance.

Another consequence of manufacturers representing their systems' typical performances with a few statistics in simple and easy to read marketing brochures and fact sheets is the difficulty users encounter when they try to compare systems from different manufacturers to determine which systems best suit their intended applications. Their decision making could be eased considerably were they privy to much of the underlying information shown

System	Accuracy	Volume Type	Volume Size
NDI Polaris	0.35 mm RMS	Pyramid Volume	1.627 m^3
Other Market Supplier	0.25 mm Mean Error	500 mm Radius Sphere	0.524 m^3

Table 1. Comparison between published specifications for the NDI Polaris position sensor and those for a competitor's system. The accuracies are for single-marker overall volume distance errors. Despite the higher value, the Polaris actually meets the competitor's specification when both systems are assessed according to a common measure.

in Figure 1. To properly define the performance of spatial measurement systems for the more "real world" types of applications that most users envision, two key items are required: (1) a sufficiently complete set of statistics that are representative of the accuracy, particularly the trueness and precision, and (2) a clear definition of the protocol that was used to generate the data from which the set of statistics was generated. Unfortunately, because of the large variety of spatial measurement technologies and expected applications, manufacturers have not agreed upon standard protocols and sets of statistics to assess their systems. For instance, it is very difficult to compare similar systems from different manufacturers in a meaningful manner if one manufacturer quotes the overall volume RMS error for an operational volume that is markedly different from that of a competitor's system, where the competitor has quoted the overall volume mean error. For example, where NDI claims that its Polaris system has an overall volume RMS error of 0.35 mm for a single marker measured at a set of specified locations throughout its pyramid (open) volume, a market supplier of a similar system quotes a mean error value of 0.25 mm over a specified spherical volume of radius 500 mm. Table 1 summarizes the specifications. The user is often expected to make an informed decision based on this limited information.

A cursory inspection of the table would suggest that the Polaris is less accurate than the competitor's system, but a more careful examination shows otherwise. The easiest comparison involves the operational volumes. The tabulated values show that the Polaris pyramid volume is about three times as large as the other spherical volume. But the shape of the volume is also important for many applications, as most applications have different spatial requirements. For example, a cranial surgical procedure in IGS may function very well in the smaller spherical volume, but an orthopedic procedure will likely require much more space. Figure 2 compares the two volumes.

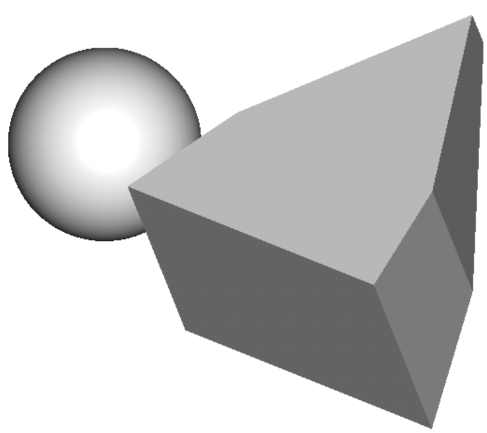

Figure 2. The NDI Polaris pyramid volume (right) and a competitor's 500 mm radius spherical volume (left).

Next, the user has to relate an RMS error taken over a larger volume to a mean error obtained from a much smaller volume. The accuracy consists of two components, the trueness, represented typically by the mean or median error, and the precision, represented typically by the variance or standard deviation. Both components are required to properly represent the accuracy, unless one component substantially dominates the other. The RMS error has the advantage of incorporating both the trueness and precision in a single value,[‡] and since distance errors are always positive, their mean will be greater than zero, implying that the RMS distance error will always be greater than the mean distance error. Using the data shown in Figure 1, the results for the two volumes can be better compared by recalculating the Polaris statistics such that only those data points falling within a 500 mm radius sphere centered at the point $(0, 0, -1900)$ are included. This amounts to more than 600 data points, which is still statistically relevant. For this sub-volume, the RMS error was 0.170 mm (compared to 0.255 mm for the full volume) and the mean error was 0.148 mm. Thus the relevant comparison should use 0.148 mm for the Polaris, which meets the competitor's specified value of 0.25 mm.

4. TYPICAL ACCURACY RESULTS FOR RIGID BODIES

NDI specifies the accuracy of its Polaris position sensor according to a single-marker volumetric calibration protocol. While this is a valid protocol for accuracy assessment, and its results have some relevance for users, it has been developed for NDI's specific needs, and so has limited usefulness for most users. Users are more interested in the accuracies they can attain with their rigid body tools, for which 6D transformations, or poses (positions and orientations) are reported. Also, for IGS applications, the 6D poses are often measured relative to an arbitrary user-defined frame of reference, which further complicates accuracy assessments. Although the single-marker results can correlate quite well with rigid body results, a proper assessment should test specific rigid bodies directly.

For hybrid Polaris systems, the rigid bodies can contain either active LEDs or passive retro-reflective spheres, and the active marker rigid bodies can have either wired or wireless configurations. Rigid bodies equipped with wired active markers can be designed according to any convenient geometry, since the markers can be easily identified unambiguously. Rigid bodies equipped with passive markers or wireless active markers differ from wired active rigid bodies in that each body being tracked simultaneously must have a unique geometry tool defintion (UGT).[6] Since these markers cannot be activated individually, the UGT definition allows for a tool's 6D poses to be determined without having to perform the initial lock-on procedure that is used to estimate the location of the markers for the prediction-and-correction tracking method. Care must be exercised when designing rigid body tools, since the tool geometry plays a very important role in its performance, and is a major source of error in IGS applications. A detailed discussion of this topic can be found the NDI Technical Bulletin by Crouch.[5] For this study, active wired rigid bodies based on a standard geometry design[6] were found to be suitable for demonstrating typical 6D accuracy results. This design has four markers located in a square configuration, approximately 50 mm apart. The markers were ceramic based IR markers that emit light nominally at 880 nm (the same marker type that is used for Polaris characterization). The passive rigid body consisted of three markers that were arranged according to UGT guidelines. Figure 3 shows the two rigid bodies. Although general studies of rigid bodies should be done on tools having geometries that are as similar as possible to allow for meaningful comparisons, the two rigid bodies that we used were sufficiently alike for the purposes of this study.

One of the questions users often ask concerns the relative accuracy obtained with tools equipped with active LED markers compared to similar tools equipped with passive markers. There is a wide-spread belief that tools equipped with passive markers cannot be tracked as accurately as tools equipped with active markers. We can examine the issue using a protocol related to the single-marker volumetric protocol used for characterization. In this case, rigid body probes consisting of either four active markers or three passive markers were stepped throughout the operational volume by the CMM for a statistically relevant number of grid points.[§] Table 2 summarizes the results, and for comparison, includes the single-marker characterization results from Figure 1, as well as single-marker calibration results from a follow-up data collection, where the reference positions differed

[‡] $RMS \approx \sqrt{\mu^2 + \sigma^2}$.

[§] The rigid body 6D poses were measured over a grid comprised of 2236 positions, of which at least 1500 were visible by both probes. At each position, 30 samples were recorded and averaged, with the averaged poses used for the data analysis. The probes were fixed rigidly to the CMM end-effector, and so their orientation was constant.

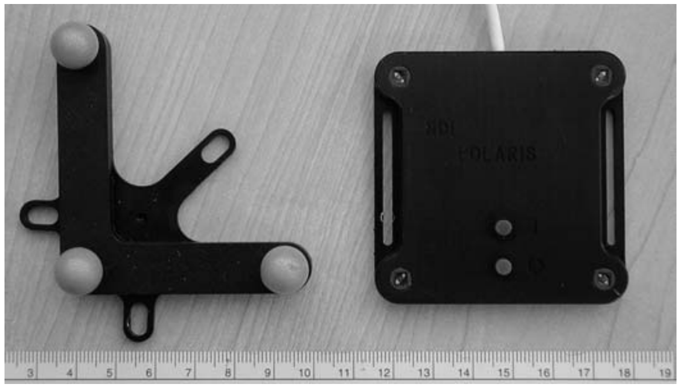

Figure 3. The passive (left) and active (right) rigid bodies used to generate the data shown in Table 2. The scale is in cm.

System	Position Errors (mm)				Orientation Errors (Degrees)		
	RMS	Mean	Std. Dev.	95% CI	Mean	Median	95% CI
Single-marker Characterization	0.255	0.193	0.167	0.451			
Single-marker Calibration	0.293	0.201	0.213	0.449			
Active Rigid Body	0.233	0.190	0.135	0.417	0.362	0.256	0.598
Passive Rigid Body	0.231	0.185	0.137	0.462	0.383	0.208	0.713

Table 2. Typical volumetric accuracies for Polaris position sensor calibrations. The characterization results are for the data shown in Figure 1. The single marker calibration results were obtained for a separate data collection done after the characterization. For the rigid body calibrations, an active four marker probe and a passive three marker probe were tracked instead of the single marker; the rigid bodies are shown in Figure 3.

from those used for the characterization, but otherwise, was similar. As with Polaris systems in general, the single-marker calibration errors are only slightly larger than the characterization errors, which implies that a follow-up calibration of the Polaris, using data collected independently of the characterization data, is not necessary, and that the characterization results themselves represent the volumetric accuracy sufficiently well (NDI does not perform additional volumetric calibration testing to verify the characterization results, but instead, uses different protocols).

For each of the rigid bodies, the position errors were defined relative to the rigid body origin, which was conveniently chosen to be simply one of the marker locations. The reference orientations for the rigid bodies were constant, and so were determined from the fits that aligned the measured grids to the reference grid. The orientation error statistics were determined according to the usual directional statistics methods.[7] As can be seen in Table 2, there is no statistically significant difference between the accuracies of the two types of rigid bodies. The rigid body results are slightly, but significantly, better than the single-marker calibration and characterization results, though, which is a consequence of the inherent averaging of the marker errors during the calculation of the 6D poses from their underlying marker positions. Generally, the rigid body accuracy improves with the number of markers whose positions are used to calculate a given 6D pose, and so the fact that the rigid body equipped with three passive spheres performed as well as the rigid body equipped with four active markers further strengthens the claim that tools equipped with passive markers can be tracked as accurately as similar tools equipped with active markers. Figure 4 shows the distance error distributions for the data sets tabulated in Table 2, and again, the passive rigid body results are very similar to the active rigid body results.

For better comparison with the active single-marker characterization results, the individual 3D marker positions were extracted from the 6D pose data for the two rigid bodies and compared directly to the reference positions. The results for each marker, as well as the average results for each marker type are shown in Table 3. The passive single-marker results are slightly worse than the active single-marker results, but not significantly

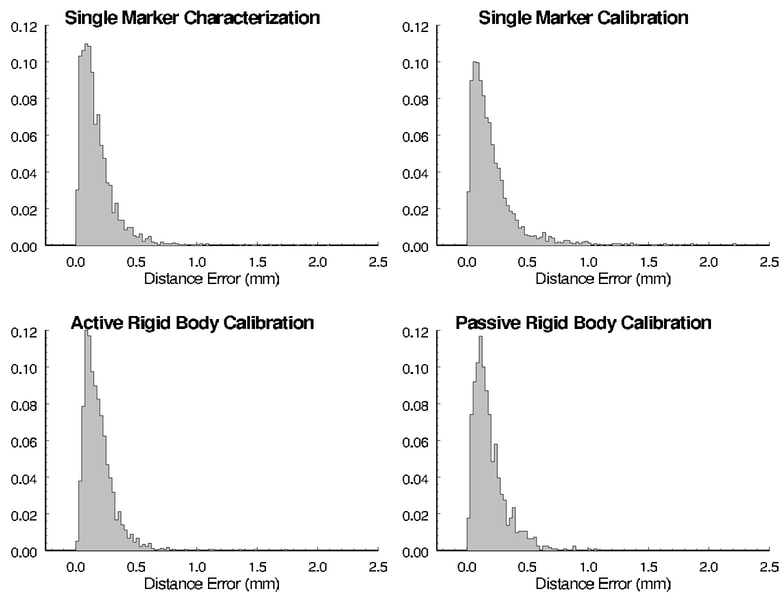

Figure 4. Frequency histograms for the distance errors shown in Table 2.

so (both are comparable to the single-marker results shown in Table 2 as well). NDI recommends that users restrict their applications to use only NDI markers to achieve the optimum accuracies, since the use of other active marker types can substantially degrade the system's performance, unless they have been thoroughly tested and proven otherwise. These results demonstrate that individual passive markers can provide similar accuracies as active markers.

Marker	RMS	Mean	Std. Dev.	95% CI
Active marker #1	0.248	0.198	0.148	0.410
Active marker #2	0.289	0.213	0.175	0.434
Active marker #3	0.273	0.210	0.175	0.409
Active marker #4	0.261	0.213	0.150	0.430
Average	0.268	0.209	0.162	0.421
Passive marker #1	0.308	0.193	0.153	0.467
Passive marker #2	0.303	0.256	0.163	0.555
Passive marker #3	0.261	0.201	0.167	0.499
Average	0.291	0.217	0.161	0.507

Table 3. Single-marker accuracies (in mm) for the individual active and passive markers, which were extracted from the rigid body results shown in Table 2.

5. APPLICATION ACCURACY

The rigid body accuracies obtained for the types of tools commonly used in IGS applications will generally not be as good as the results shown in Table 2, since these tools typically have their origins defined at their probe tips, which can be located several cm from their markers. Small errors in the marker positions may tend to be averaged out during the calculation of the 6D poses, but the residual errors can magnify substantially when they are extrapolated to the tool tip. This can be seen in Figure 5, which shows how the rigid body accuracy can degrade when the distance between the markers and the rigid body origin increases. The accuracy is related to the orientational uncertainty as

$$\epsilon_i \propto d_{\mathrm{tip}} \cdot \tan \delta_i, \tag{4}$$

where ϵ_i is the error at the tool tip for measurement i, d_{tip} is the distance from the rigid body markers to the tool tip, and δ_i is the orientational uncertainty for the given measurement. The data were obtained from the volumetric protocol discussed in Section 4 for the active marker rigid body. For those data, the origin of the rigid body was simply located at one of the four markers. For this assessment, four more rigid body definitions were created with the origin shifted 25, 50, 75 and 100 mm from the original origin in the rigid body's z-direction. The accuracy statistics were reprocessed according to these four new rigid body definitions. Since the orientation of the body is independent of the rigid body origin, the orientation uncertainty δ_i was taken to be constant. As can be seen in Figure 5, the dependence of the error on the tool tip distance is substantial, with the RMS error doubling over the 100 mm range, from 0.233 mm to 0.470 mm.

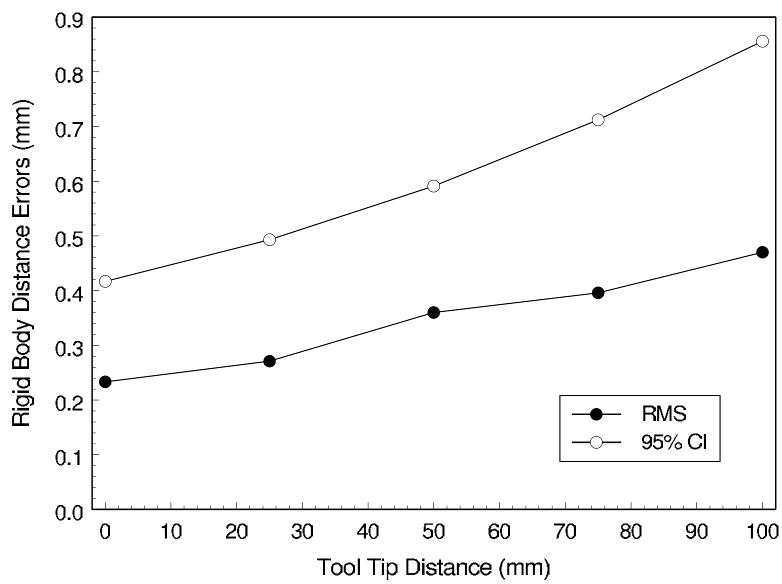

Figure 5. Inherent accuracy degradation of rigid body 3D positions as the tool tip distance is increased with respect to the rigid body markers. The RMS and 95% confidence interval accuracy results are presented for five different tool definitions. Each definition is identical, except that the tool tip position has been redefined to be at a different distance. The resulting accuracy values are plotted against this virtual tool tip length. The data were obtained from the active rigid body data shown in Table 2.

Taking into account the effects of the probe tool tip distance helps make the protocol more useful, but the results shown in Figure 5 and in Table 2 were all obtained with the rigid bodies held fixed with constant orientations, so that many marker characteristics were not an important factor. This would not be the case for probes used in typical IGS applications. So, while this volumetric calibration protocol was good enough for demonstrating that passive tools have accuracies comparable to active tools of similar design, it is too limited to serve as an assessment for general IGS applications. Also, very few users have access to a CMM, which makes the protocol too inconvenient for most users.

Spatial tracking systems used in the field can be affected by rough handling and have their accuracy degrade. Thus it is important that users check their systems periodically with an appropriate calibration program to ensure that they are performing to the levels needed by their applications. A calibration program is a set of procedures undertaken periodically to determine whether a given measurement system needs to be returned to the manufacturer, or to some qualified third party group, to be readjusted to meet the required measurement performance. These procedures include a calibration test, or set of tests, as well as a set of criteria used to determine whether the system passes. Given the extensive number of different applications making use of tracking systems, manufacturers cannot develop single protocols that would be relevant for all users, and so users should develop their own calibration protocols, preferably ones that are intimately related to their usual applications. Users must also determine their own pass/fail criteria, since the measurement accuracy required by one application might be quite different from that required by another application, and neither might meet any of the criteria specified by manufacturers.

Designing application-specific accuracy assessment protocols must be done carefully, as there are numerous factors to be considered, many of them not at all obvious. Some of the important factors to consider include the size of the calibration volume (few applications make use of the entire operational volume), the rigid body design for tracked tools (ideally, the calibration protocol would use tools very similar to those used in the application), the tool orientations (the protocol should include a representative sampling of the orientations typically encountered by the application), the point of interest on the tracked tools (typically, the tool tip), and whether a local reference frame is used. For example, Frantz et al.[2] discuss the use of a device for calibrating tracking systems based on electromagnetic technology. The device is a hemispheric artefact that has 50 slots randomly situated on its surface that can each accommodate, in a repeatable manner, standard rigid body probes equipped with 5D or 6D magnetic sensors. The hemisphere is roughly the size of a human skull, which makes it suitable for use as a phantom in a cranial IGS procedure.

6. CONCLUSIONS

When manufacturers characterize their spatial measurement systems, they choose procedures that are best suited for generating the model parameters needed to convert the systems' raw signals to their corresponding measured positions. These procedures usually involve volumetric protocols that can be used as calibrations as well, and such calibrations can provide one assessment of a system's accuracy. Such assessments are often represented in marketing material as the system's general "accuracy," but such specifications are of limited use for users, since the procedures best suited for characterization seldom encompass the more general system usage that users demand for their applications. Also, the few "representative" statistics typically presented to users do not contain much of the important underlying information often required by users to properly assess a given system's suitability for their intended applications. Manufacturers often use these statistics to their advantage to enhance the perceived performance of their systems in comparison to their competitors' systems by selectively presenting statistics having inherently lower values. Users considering measurement systems for purchase must be very careful when examining representative statistics from different manufacturers to ensure that both the statistical values and the calibration protocols on which they are based are indeed comparable.

Manufacturers of optical tracking systems that track rigid bodies often represent their systems' accuracies with single-marker results instead of rigid-body results, which would be more relevant for IGS applications. Accuracy assessments for systems tracking rigid bodies are much more ambiguous, since the specific rigid body being used has a large effect on the perceived accuracy. For a given rigid body, the accuracy that can be achieved depends on many factors. In general, increasing the number of markers increases the accuracy, because of the inherent averaging of individual marker position errors when the corresponding 6D poses are determined, but this increase is typically offset more by the decrease in accuracy resulting from the tool tips of most probes used in IGS applications being located several cm from the rigid body markers. Since most applications use different types of rigid bodies, "typical" rigid body accuracies cannot be specified, and so the single-marker specifications have the advantage of serving as a common baseline for comparisons, despite the paucity of single-marker applications. For the NDI Polaris position sensor, tools equipped with passive markers can be tracked as accurately as similar tools equipped with active markers, contrary to beliefs held by many users.

ACKNOWLEDGMENTS

Special thanks go to Stephen Leis, who helped with some of the early experimental design, and to Kristian Dixon who helped with some of the figures.

REFERENCES

1. P. Kantrowitz, G. Kousourou, and L. Zucker, *Electronic Measurements*, Prentice Hall, Inc., Englewood Cliffs, NJ, 1979.

2. D. D. Frantz, A. D. Wiles, S. E. Leis, and S. R. Kirsch, "Accuracy assessment protocols for electromagnetic tracking systems," *Phys. Med. Biol* **48**, pp. 2241–2251, 2003.
3. S. E. Leis, "NDI-TB-0005: Polaris calibration performance and methodology, Rev. 002," tech. rep., Northern Digital Inc., 1996.
4. D. Hsu, *Spatial Error Analysis, A Unified Application–Oriented Treatment*, IEEE Press, New York, 1999.
5. D. G. Crouch, "NDI-TB-0021: Design and manufacturing tools incorporating ired markers, Rev. 002," tech. rep., Northern Digital Inc., 1995.
6. Northern Digital Inc., "NDI-TB-0004: Unique geometry specifications for polaris system tools, Rev. 009," tech. rep., Northern Digital Inc., 2003.
7. K. V. Mardia and P. E. Jupp, *Directional Statistics*, John Wiley & Sons, Ltd., New York, 2000.