

Management of Tracking for Industrial AR Setups

Peter Keitler*
Technische Universität
München

Benjamin Becker†
EADS Innovation Works

Gudrun Klinker‡
Technische Universität
München

ABSTRACT

The accuracy of a real time tracking system for industrial AR (IAR) applications often needs to comply with production tolerances. Such a system typically incorporates different off-/online devices so that the overall precision and accuracy cannot be trivially stated. Additionally, tracking needs to be flexible to not interfere with existing working processes and it needs to be operated and maintained free of error by on-site personnel who typically have a quality management (QM) background. For the final validation of such a complex tracking setup, empiric testing alone is either too expensive or lacks generality.

This paper demonstrates a new approach to define and verify, deploy and validate, as well as to operate and maintain an IAR tracking infrastructure. We develop our concepts on the basis of an IAR application in the field of QM in the aircraft production process. It integrates a qualitative visual comparison with accurate quantitative measurements of 3D coordinates using a metrological probe. The focus is on the verification, validation, and error free operation.

Monte Carlo simulation predicts the error for arbitrary system states. Using a limited set of empiric measurements in the target environment allows us to validate the simulation and thereby validate the application. This combination assures compliance of the IAR application with the required production tolerances.

We show that our simulation model yields realistic results, using an in-depth analysis of an optical IR tracking system and a high-precision coordinate measurement machine capable of densely sampling the entire tracking volume. Additionally, it allows for a straightforward derivation of run-time consistency checks for the automatic identification of possible system failures. Also, estimation of the system performance during the planning and definition phases becomes possible, using the elementary accuracy specifications of the involved sensor systems.

1 MOTIVATION

A tracking infrastructure is required in all industrial AR (IAR) applications to perceive and interact simultaneously with real and virtual objects. Especially in the naval and avionic industry, products are very complex and highly customized. The high rate of manual work is an opportunity to create new tools using AR technology that have the potential to decrease production time and cost. Thereby, the reliability, robustness, and accuracy of the tracking infrastructure is an important factor for the success of such applications. Besides the tracking system, there are various other application constraints towards usability and work safety that have to be dealt with, but which are out of the scope of this paper. A detailed description of the challenge to bring AR out of the laboratory into an industrial context is provided in [6].

An exemplary IAR application for the support of quality management (QM) processes in the avionic industry is depicted in Fig-



Figure 1: Industrial Augmented Reality



Figure 2: Metrological Measurement & Tracking Device

ure 1, where a worker sees an x-ray view on the electric wiring. Besides qualitative comparison by video overlay or display of purely virtual content, there is a need for quantitative measurements in real time, e.g. by using a probing device as depicted in Figure 2. Therefore, a certain level of accuracy has to be guaranteed for all specified modes of operation. The purpose is to use AR as a normal tool, fully integrated into the working process. However, many factors might complicate the design of a suitable tracking infrastructure. Tools within the existing work procedures in the target environment need to be robustly tracked using added markers. Also for registration of the virtual and the real model, markers have to be attached to the product itself. Besides typical issues like occlusion and reliability, such modifications are strictly limited in order not to compromise the existing process and product.

Especially in the avionics industry, another problem is the large potential area the tracking system should be used in. This requires solutions to increase the flexibility of the tracking system. Manufacturing procedures might require removal and reinstall of the tracking system, invalidating the registration and resulting in recalibration and maintenance effort. Ideally, such procedures can be performed through on-site personnel without requiring detailed

*e-mail: keitler@in.tum.de

†e-mail: benjamin.becker@eads.net

‡e-mail: klinker@in.tum.de

technological insight by providing an appropriate user interface.

In most common IAR scenarios, these issues exist at least partially. Often, they can only be addressed by strongly limiting the use case scenario or by the integration of various sensor devices. Especially for those complex IAR setups it is challenging to assure compliance with requirements like production tolerance. This has to cover the entire error chain, consisting of the individual sensor errors accumulated and propagated throughout the entire system setup. This therefore covers not only the real-time error but also registration procedures of rigidly mounted sensors, objects, and markers, based on error-prone measurements [10]. Typical generic vendor specifications are often insufficient to provide a reliable assessment of the overall tracking system performance in the explicit use case scenario.

The goal of this paper is not to find solutions to all these IAR problems¹. Many solutions exist in the literature to provide pragmatic solutions to special problems, e.g. [9][14]. Rather, general guidelines shall be provided to handle the complexity of the tracking setup, with a focus on the need in IAR for a well-defined tracking accuracy.

Following classic approaches from the metrological domain in production engineering, the validation of the tracking infrastructure would have to be accomplished through exhaustive empiric measurements, see e.g. [22][23]. This is affordable only for single systems such as a mechanical measurement arm or a theodolite that can easily be analyzed in a controlled setup. A complex tracking infrastructure incorporates several sensor systems that potentially have to be rigidly installed in the target environment, making the acquisition of empiric measurements for comparison very difficult.

In this paper we propose to combine a small set of representative empiric measurements with a comprehensive simulation to cover all possible constellations and system configurations in the use case scenario. We rely on a tracking framework to model the simulation using standardized and integrated components for all necessary computations. This approach requires identifying the degree to which the simulation model is an accurate representation of the real world [2].

After introducing the relevant literature in this field, Section 3 describes our approach for the various design and implementation stages of an IAR application. Section 4 details our exemplary IAR scenario depicted in Figure 1&2. Section 5 describes in detail the application of our approach which is then evaluated in Section 6 on the basis of our exemplary scenario. It will be shown that simulation methods scale well to such complex setups and provide reasonable and valid results.

2 RELATED WORK

Prior work to analyze the accuracy of tracking setups and systems can mainly be categorized in online and offline error estimation/propagation techniques and the analysis of individual systems.

Online error propagation approaches such as [3][5][9] provide an estimate of the error in real time. It is based on the current system state, e.g. distances between camera and marker as well as the visibility of fiducials. Due to performance reasons, the prediction throughout the entire error propagation chain usually requires the linearization of each computational step. This simplification is fully sufficient to give the user an real-time insight into the current system performance. In IAR however, the validation of a tracking system requires an assessment of the estimated overall tracking accuracy before deployment.

Offline techniques are split into two domains, simulation and system analysis using ground truth measurements. Simulation techniques have been applied successfully to assess the accuracy of offline photogrammetric methods or metrological measurement de-

VICES such as theodolites or laser trackers as well as real time tracking systems. The goal is to validate a system or to detect the main sources of error and to point out optimization potentials. Pentrieder et al. use a Monte Carlo simulation approach to predict the accuracy of an optical square marker tracker by automatic generation of artificial camera images [17]. Hastedt also uses a Monte Carlo approach to simulate the behavior of a photogrammetric system [8]. Both validate the simulation approach using high-precision measurement data. Simulation has also been used to assess algorithm performance [4][7].

'Ground truth' measurements have also been used many times to benchmark various types of tracking systems. Schmidt et al. use a high precision linear stage to evaluate various IR tracking systems manufactured by NDI² along the three spatial axes [21]. Satoh et al. use an industrial robot to move an HMD through the tracking volume to evaluate the VICON³ and LaserBird⁴ optical tracking systems [20]. They provide a framework which is applicable to various systems but requires dense ground truth data. However, the practical use of such a framework is limited since it requires a robot to provide dense empirical measurement data, which is not feasible in a typical IAR target environment, as described in Section 1. Also, no general statement for design criteria such as the world registration procedure or the layout of a marker, which have a strong impact on the overall accuracy, is possible. Although an elementary uncertainty specification is determined, it can hardly be reused because typically too many factors will differ, even in similar scenarios. This is also covered by our own experience with optical IR tracking [14]. Other works restrict themselves to the assessment of an individual system. Lieberknecht et al. use a FARO measurement arm to generate ground truth image data which is then used to assess the quality of different markerless tracking algorithms [15]. Similar work has been done in the medical domain where Rohling benchmarked the tracking of medical instruments using a FARO measurement system as a ground truth [19]. Even vendors provide insight into their systems to propose means to detect invalid calibration [23].

Allen et al. presented a general method for the evaluation and comparison of the expected performance of tracking systems [1]. Their approach even incorporates scene dynamics and provides advanced visualization concepts. However, their simulation considers only a single tracking system using a moving target and a pose estimation algorithm. The required registration and calibration procedures within IAR scenarios are not covered.

3 APPROACH

It is important to provide means to setup and implement robust and maintainable tracking infrastructures to allow AR applications to spread in the industrial domain. We already proposed the graphical tool *trackman* for this purpose [13]. It is based on the *Ubi-track*⁵ library. The spatial relationship graph (SRG) and spatial relationship pattern concept allows for the creation of efficient and semantically correct runtime data flow descriptions for arbitrary calibration/registration and runtime tracking problems [18]. This generic approach shall now be strengthened, to obtain an integrated approach for the management of tracking in industrial setups.

Three important development phases can be distinguished in an industrial context. During the *definition* phase, the tracking system is planned, algorithms are implemented or combined, and buying decisions are taken. Then, the planned system is installed in the target environment. This *deployment* phase also incorporates the initial calibration of devices and registration of spatial transformations

²www.ndigital.com

³www.vicon.com

⁴www.ascension-tech.com

⁵campar.in.tum.de/UbiTrack/WebHome

¹although we had to tackle some of the described issues in our exemplary scenario, of course

between sensors, markers, and objects. Subsequently, the system is used productively in the *operation and maintenance* phase.

As displayed in Figure 3, we extend the definition phase by simulation based verification, the deployment with a validation step and the final tracking system operation phase by run-time checks that detect system malfunction and propose mitigation steps to the user.

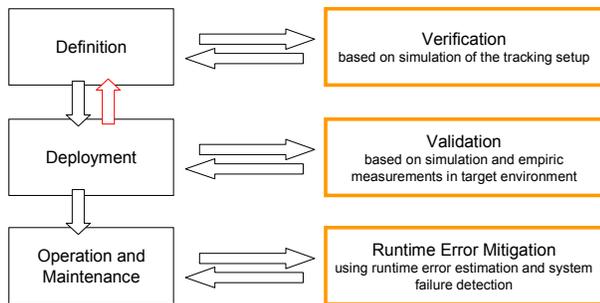


Figure 3: Proposed development process for an industrial tracking infrastructure

To verify the design we propose to not only rely on analyzing the stand-alone tracking system in detail but to expand the evaluation to the entire error chain of the IAR application. Using the *elementary uncertainty specifications* of the involved sensor systems in a Monte Carlo based simulation, it is possible to derive the behavior of the complete setup. Since the simulation approach requires known uncertainties of the used tracking devices, the analytic evaluation as presented in Section 2 is of great importance. Using our simulation it is possible to benchmark various hypothetical scenarios without needing real hardware. Thereby, it can help to state the general feasibility of a setup, to justify buying decisions, and to decide on which particular algorithms to use. The final goal of the simulation is to verify the overall concept and to decide on a certain variant before deployment.

Generally, the correctness of a simulation system for a certain purpose first needs to be proven [2]. This assures that the assumptions made are specific enough to obtain realistic results. Based on our use case, we demonstrate that existing simulation approaches used to evaluate individual tracking systems and algorithms (cf. Section 2) can be extended to a more complex IAR tracking setup.

Following industrial standards, classical validation would require exhaustive empirical measurements to validate the entire system operation [2][22]. As shown in Figure 3, we reduce the effort using the simulation to validating that the simulation represents a realistic system behavior. If the assumptions for the simulations are incorrect for the productive environment, a redefinition of the system might be required, resulting in an iteration with the definition phase (red arrow).

As proposed in Figure 3, such detailed system knowledge allows to implement runtime error mitigation using consistency checks for the automatic identification of possible system malfunction. These checks are based on the detection of violations of systematic and random error limits that were developed during the definition and deployment stages. Those can be interpreted and traced back to a certain system component and propose mitigation steps to the user.

To allow for Monte Carlo simulation of a planned tracking setup, an *elementary uncertainty specification* is needed for all involved sensors:

Abstraction Level: As input, our simulation framework uses sensor uncertainty specifications at different abstraction levels. Depending on the used system, it can be given for the individual 2D measurements of a camera, for the 3D positions of fiducials, or for the 6DoF pose of a marker. A simple isotropic error model or a more general covariance matrix are supported. The decision on

simulation system granularity highly depends on the used system setup. If the layout and the amount of fiducials allowed for marker design are predefined, it could make sense to specify a 6DoF pose error. Similarly, in an optical multi-camera tracking setup with arbitrary camera arrangement, rather a 2D value is specified since the uncertainty of a 3D position of a fiducial depends on the amount, distances, and distribution of cameras. Sensor uncertainties can and should be specified at a level that allows for maximum generality. This concerns specifications provided by system vendors as well as third-party accuracy assessments (cf. Section 2).

Error Distribution: According to the *Guide to the Expression of Uncertainty in Measurement (GUM)* [12], sensor noise may follow different error distributions. By default, one might assume a Gaussian error distribution. However, there are tracking systems where global systematic distortions of the tracking volume dominate the overall error behavior [14]. In such cases, a uniform distribution could be more realistic in order not to underestimate extreme values. Also a combination of different error distribution might be used since systems often suffer from sensor noise as well as systematic errors.

Error Magnitude: Finally the magnitude of the error has to be provided. Again, this can be either taken from the detailed system analysis or from the system vendor specification. Also the degree of the noise depends on the setup and on environmental influences. If the actual amount of error is not known a priori, simulation with different assumed error levels can still help to define the maximum allowed sensor noise to stay in-line with the overall application specifications.

4 USE CASE SCENARIO

A typical QM user needs to qualitatively compare the digital and the physical product using a visualization device (see Figure 1). Due to the size and complexity of the aircraft, there is a strong benefit in tracking this device to offer a real-time view on the digital model. Additionally, it is of great importance to assess production quality by quantitatively validating the location of new parts through length or coordinate measurements. This is usually done using a metrological device (see Figure 2). Both are integrated into an IAR application which is used integrated in the production processes.

Already experts use metrology systems to provide measurements within the production environment. For these precise offline measurements within a production environment as large as our aircraft, photogrammetric or laser based metrology systems are used. Such systems are less easily integrated in the standard production processes and often require a fixed registration sensitive to vibrations and therefore interrupt other production process running in parallel.

To cover the large area of the aircraft where the QM procedures are performed, it is not feasible to deploy countless tracking systems. Not only would this result in high cost but also would the permanent installation of tracking hardware collide with other production processes. Rather a quick setup and dismantling procedure is required. We therefore use a mobile tracking system that is able to reference itself within the aircraft and to perform the real time tracking of the visualization and probing devices. This is done by adding reference targets to the aircraft. Their static transformation within the aircraft's coordinate frame is determined in a registration routine using an offline metrological system. Therefore, this could be called an *indirect tracking setup*, see also [14].

For application in the QM process, it is fundamental to assess whether the accuracy of the tracking setup is in line with the production tolerances. This requires assessing the accuracy of the tracking system throughout the entire error chain from registration to real time tracking. Evaluation using ground truth measurements in the target environment is unfeasible due to the huge operating volume paired with varying situations and conditions. Therefore, we apply our approach using simulation throughout the development phases.

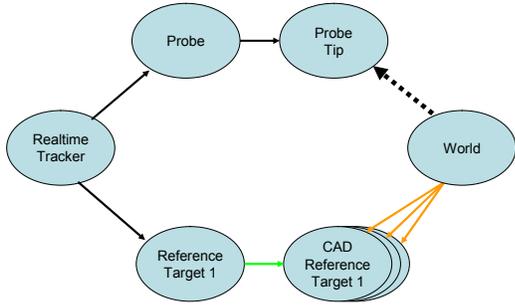


Figure 4: Use case spatial relationship graph (SRG)

Figure 4 shows the spatial relations between the involved entities. Nodes represent coordinate frames, edges depict spatial transformations. Initially, the reference targets are added to the aircraft. They are precisely manufactured and can be detected by the offline metrological system. Measuring multiple points on the CAD model of the reference target allows computing its 6DoF rigid transformation in the aircraft (World→Cad-Reference-Target) using an absolute orientation algorithm, e.g. [11]. This measurement is prone to the error of the offline metrological system.

Second, we add a real time IR tracking system which can identify and estimate the 6DoF pose of LED targets such as shown in Figure 2. The system computes the pose of the targets internally using an unknown absolute orientation algorithm (Real Time-Tracker→Probe, Real Time-Tracker→Reference-Target). These pose transformations are subject to error due to the positional error of each LED. Additionally, to use the probing device, a tip calibration of the 3D offset between the LED target and the tip is required. Using many samples leaves a negligibly small error on this calibration routine.

The final missing registration is between the CAD model and the LED marker of the reference marker. This is done by measuring 25 known reference points⁶ on the CAD model with the probe and computing the transformation using an absolute orientation algorithm (Reference-Target→Cad-Reference-Target, see Figure 5).

Completing the required registrations allows to compute the position of the probe tip in the world (dashed edge) - the basic for our IAR application. This is derived by applying trivial inversions and concatenations to the transformations in the depicted SRG [18][13]. The tablet PC for qualitative visualization has been omitted for simplicity, it can be tracked analogously to the probe.

5 CONCEPT

This section explains the application of our simulation-based concept for verification, validation, and runtime error mitigation depicted in Figure 3. The results, as well as the correctness of our approach, will be addressed in Section 6.

5.1 Verification

During the definition phase, the proposed system setup shall be checked for compliance with the application requirements, by means of a suitable verification process. First of all, the application data flow has to be modeled on the basis of the SRG depicted in Figure 4. The application SRG serves as a basis for the data flow descriptions to be instantiated using the tracking framework [13].

Since the simulation has to cover the entire error chain, also the registration data flows are required. They are obtained by a refinement of the application SRG (cf. Figure 4). In our case the level of abstraction requires to explicitly formulate the algorithms that are

⁶A large number is used to mitigate the influence of probing errors.

actually performed inside the black-box tracking system to use the sensor accuracy description that is given for a single LEDs only.

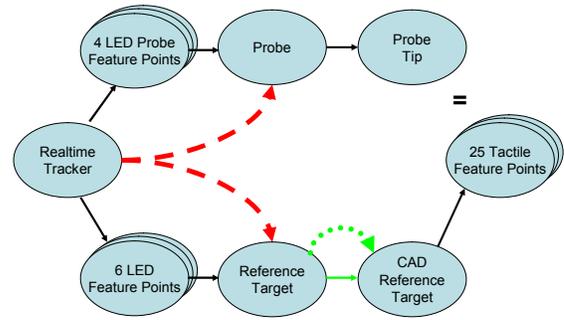


Figure 5: Refined SRG

Figure 5 depicts this situation. There are two absolute orientations that are performed internally in the real time tracking device locating the LED features in its coordinate frame (dashed edge).

The calibration routine to derive the transformation between the reference target used by the real time tracking and its CAD model uses the CAD features measured by the tip of the probe (Tactile-Feature-Points=Probe-Tip). The cycle in the SRG allows computing the dotted transformation using an absolute orientation.

Simulation Preconditions For the simulation of the entire error chain, all registration procedures as well as the final application data flow have to be simulated using the elementary uncertainty specifications, i.e. the abstraction level as well as the distribution and magnitude of the errors (cf. Section 3). In our case this is less complex since the system is pre-calibrated and uses three rigidly installed cameras (see Figure 2).

The system provides 6DoF poses of the probe and reference targets. Nevertheless it is not feasible to specify the error at that abstraction level since we use custom marker layouts. The error of 6DoF pose tracking highly depends on layout and dimension and therefore cannot be stated in a general way by the vendor. Applying the error to the 3D positions of the individual LEDs is appropriate since this is the provided elementary uncertainty specification⁷ for the position of a tracked LED. An uncertainty specification on the sensor's image plane using 2D noise would also be possible, however internal details about the intrinsic and extrinsic camera parameters would be required for this. It is apparent that the used system and setup define the used granularity.

For our use case we decided to use a non-isotropic Gaussian error distribution to approximate the real error distribution, following the vendor's elementary uncertainty specification with an overweight error in the depth direction. This might not be a perfect assumption for the complex systematic error we have found in our in-depth analysis but it is a sufficient bound.

Simulation Setup The simulation was divided into two parts: in the first step, static spatial transformations are estimated (registration), then the overall application accuracy is estimated (operation) in a second step. During operation, all pre-calibrated transformations (colored in Figure 4) from prior registration are used, together with their estimated covariances.

Ground Truth Data: The simulation system shall imitate the expected constellations of the planned real system. For this, hypothetical data has to be provided in terms of assumed spatial relationships. This also incorporates static transformations that might be known from preceding simulation steps. This synthetic data represents a ground truth; it does not contain any measurement errors. In

⁷www.ndigital.com/industrial/certushd.php, RMS error [mm] for horizontal/vertical/depth directions (depending on depth): 0.1/0.1/0.15[mm] (at 2m), 0.15/0.15/0.25[mm] (at 4m), 0.25/0.25/0.45[mm] (at 6m)

our scenario this includes exemplary poses for the reference target in the airplane, possible poses of the real time tracker with respect to the reference target, as well as an exemplary grid of probe poses in the volume of the real time tracker for which the uncertainties will be simulated. Also the layouts of tracking targets are specified, in our case defining the LED constellations. They can typically be obtained directly from the proprietary target calibration routines provided by the system vendor.

Synthetic Measurements: Based on the ground truth data, synthetic measurements are generated. To apply the sensor noise, the ground truth data needs to be transformed into the coordinate system(s) in which the elementary sensor uncertainty specification is provided. By sampling from the given probability density function the data is perturbed accordingly. Using the estimated covariance, noise is also applied to the static transformations from precedent simulation steps. In our case, the LED positions in the coordinate system of the real time tracker have to be derived from the assumed 6DoF poses of the targets and the relative 3D offsets of the LEDs belonging to these targets. Another example would be to transform 3D points to the image plane of a camera and apply 2D pixel noise there, if the elementary uncertainty specification were given at that level.

Registration/Tracking Algorithm: Now, the perturbed spatial transformations are propagated through the SRG to the coordinate frame in which they are required for the intended calibration/registration or application. This represents the normal operation of a Ubitrack data flow. Various kinds of spatial relationship patterns are at our disposal, besides the trivial inversion and multiplication, this comprises many common calibration and registration methods. The absolute orientation pattern using 3D-3D point correspondences is just one example, see also [18]. In the last simulation step of our evaluation, the tip of the probing device is estimated in world coordinates. Similarly, one could also estimate the 2D overlay error on an HMD by propagating the error to its image plane.

Covariance Estimation: Now, the previous two steps are repeated iteratively. Samples are produced by perturbation and then propagated to the coordinate frame of interest by running the corresponding algorithms. Accumulating those, using descriptive statistics, the covariance associated with the registration or tracking result can be estimated [16][12].

To better understand this approach, we give an exemplary description of estimating the covariance of the registration of the reference target using 25 tactile points, based on our refined SRG depicted in Figure 5. Our synthetic data describes that the tactile points will be probed consecutively. Also, the real time tracking system is positioned in the setup so that the reference target is at an optimal distance with minimal error.

To apply the noise in our setup, we use the ground truth of the probe keeping the tip fixed on the one of the tactile points. We derive the location of the LEDs of the probe marker and the reference target in the coordinate frame of the real time tracking. There, the Gaussian noise is applied to each LED position. Using an absolute orientation algorithm, an error-prone estimate for the probe and the reference target and consequently for the tactile point and the probe tip is computed. Iterating over the tactile points, we derive 25 corresponding 3D point pairs (inner loop). The erroneous registration of the LEDs with the tactile points is obtained by another absolute orientation. Sampling this transformation multiple times results in the desired registration and the associated covariance (outer loop).

The computed registration and covariance is then used in the subsequent simulation step for the entire system. It follows the same principle. We specify several poses of the real time tracker with respect to the reference target, as well as grid of assumed probe positions in the tracking volume. Again the simulation iterates over these poses, perturbing the pre-calibrated transformations as well as of each of the currently visible LED positions to sample the covari-

ances of the probe at the different positions in the tracking volume.

5.2 Validation

Following the ASME standard [2], validation is 'the process of determining the degree to which a model is an accurate representation of the real world. The intention is to validate the simulation and its preconditions by performing some selected experiments in the target environment under varying conditions. This has the character of a mandatory final inspection of the system after deployment, to show that all relevant influences have actually been considered and the defined system behaves as expected in the target environment. This validation is *not* suited for the derivation of a suitable elementary sensor noise specification and corresponding simulation model based on it. It might however help to estimate the error magnitude, assuming that abstraction level and error distribution are known (cf. Section 3). For now we assume that the chosen preconditions and simulation model are basically correct; a proof will be given in 6.1, based on extensive empiric measurements in a controlled setup.

In our case the simulation predicts the position of the tip at a certain position in the world. For comparison, a corresponding set of real reference points in world coordinates is needed, known with high accuracy. This explains why validation is expensive and should be reduced to the necessary minimum. In our case, this ground truth is obtained from the multiple reference targets whose tactile points have already been registered with the world using a metrological device. While one target is being used for indirect tracking, points on this and also other targets can be probed.

Depending on the complexity of the setup, a step-by-step validation might be more efficient. First, the assumed elementary uncertainty specification is validated in the real environment, based on a few reference measurements. This might reveal potential system malfunctions due to so far unconsidered environmental influences or invalidated system calibration. Next, the individual simulation steps for registration (cf. 5.1) are validated. The procedure is carried out several times under varying conditions (distance, viewing angle, ...) and a covariance is estimated from the individual results. In our case, this applies to the registration procedure for the reference target described above (cf. Figure 5, dotted edge). Differences between the simulated and empiric covariances could point to user errors such as point mismatches or imprecise probing. Finally, the complete error propagation chain is validated (see above). When these steps are completed, one can assume the simulation represents the real world sufficiently. This allows using the simulation as the basis for verification & validation of the entire system.

Statistical methods can be applied to formally reject the hypothesis of contradictory measurements and covariances. It can be tested whether all individual empiric measurements reside inside the predicted confidence intervals. A *T-test* decides whether the means of two data sets are identical, e.g. when a series of measurements of a certain point or pose shall be compared with the predicted mean and covariance. It can reveal systematic errors. Furthermore, the identity of a predicted with an empiric covariance matrix can be shown using the *Box' M-test*. In a static setup this indicates that the elementary uncertainty specification used for simulation may be incorrect. In a dynamic setup, as for example a probe rotating on its tip, it indicates an error in the calibration of the probe or its tip.

Validation showed that the simulation describes the real tracking infrastructure sufficiently. As described in Section 4, this is fundamental for the qualification of industrial processes. Furthermore, the findings can also help to maintain the proper condition of the infrastructure during runtime, as described next.

5.3 Runtime Error Mitigation

During operation & maintenance, means are needed to reliably detect, trace back, and eliminate potential system failures. The expected accuracy and precision have been analyzed through the ver-

ification and validation procedures. This allows us to define constraints within the tracking setup’s SRG to compare current measurements with the expected system behavior at run time.

The simplest way to perform a run time test is a ground truth measurement. When probing a known point it is possible to compare the measurement to the expected value. Using the same statistical tests as during validation (cf. 5.2), it is possible to determine whether a measured point set violates the expected confidence interval. In case the user can reproduce the effect, it is not caused through a user error like imprecise probing as for example slipping or measuring the incorrect point. This indicates a malfunction. Due to the complexity of current tracking setups such an error cause could not be easily traced back.

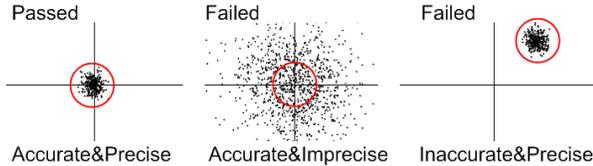


Figure 6: Runtime checks, ○ represents the confidence interval.

To locate the error source, it is possible to run the validation tests simply in reverse order. The first step validates the probe tracking in the complete system. Measuring with the probe tip on a known fixed point (ground truth) while rotating the probe results in a position and a covariance ellipsoid of the tip in the world.

The resulting measurement can violate the specified error level in two ways. In case the covariance is larger than the simulation predicts, the source of error should be in the direct probe tracking (Figure 6 accurate & imprecise). Recalibration of the probe target and the tip calibration should solve that issue. Otherwise, the error has the correct magnitude of noise but the measured position is incorrect (Figure 6: inaccurate & precise).

In the latter case a second test needs to be performed by probing at least 3 known tactile points on the reference target and comparing the measurement to the ground truth values from the CAD model in the coordinate frame of the reference target. If these points are not within the confidence interval, this indicates that the transformation between the 25 tactile points and the marker of the reference plate is invalid. Recalibration of the reference target solves this issue.

If the test is passed and the points are correctly mapped, the only left registration is between the reference marker and the aircraft (world). A recalibration of this reference using the offline metrological system is necessary.

After this procedure, the final system test is repeated by measuring a known point in the aircraft. In case the distance between the measured point and the ground truth exceeds the confidence interval, there are two severe error sources left. Either the offline metrological system was not registered properly to the aircraft in the calibration procedure or the real time tracking system is outside its specification.

Performing this procedure at runtime asserts that the tracking system and the work procedures relying on the new application are in line with the specification. In a typical scenario such an invalid state should not occur often or the tracking setup should be revised.

6 EVALUATION AND RESULTS

The evaluation of our approach is three-fold: First, we prove that the simulation of the indirect tracking scenario yields reasonable results. For this, the consistency of simulation and empiric data is shown based on extensive measurements in a controlled setup. This step will not be required for comparable setups in the future. Next, we verify the expected performance of our application in the target

environment, based on exhaustive simulation. Finally, we validate our application in the target environment, based on some selected measurements.

6.1 Consistency of Simulation



Figure 7: Test-Target for Various Metrological Systems

To evaluate the consistency of the simulation, we rely on distance measurements as proposed by international metrological standards [22, 2], as well as point-based registration and adjustment theory [16]. We designed a test target (small picture in Figure 7) that consists of a tracking target for the online metrological system⁸ (black LED target) and a laser reflection target (silver sphere) for the offline metrological system, a high-precision laser tracker⁹. Both tracking systems are located in front of the target to be able to identify the reference target’s positions.

A high precision coordinate measurement machine¹⁰ (CMM) is used as a reference. It allows us to move the test target to an arbitrary specified location within a volume of $6 \times 4 \times 2.5[m]$, with an accuracy of $10\mu m + 14\mu m/m$ from its coordinate origin. We programmed the CMM to scan the entire measurement volume as shown by the small spheres in Figure 8, stopping at each position for 3 seconds. The distance between the grid positions was $20cm$ in each direction. We furthermore scanned three orthogonal lines along the coordinate axes of the real time tracker, at a regular distance of $1cm$. The frustum of pyramid in Figure 8 indicates the tracking volume as specified by the vendor.

Pausing at every position for 3 seconds allowed us to synchronize programmatically the three systems afterwards, as well as to assess the precision (affected by noise) and accuracy (affected by

⁸NDI Optotrak, www.ndigital.com

⁹FARO Ion, www.faro.com

¹⁰DEA Lambda

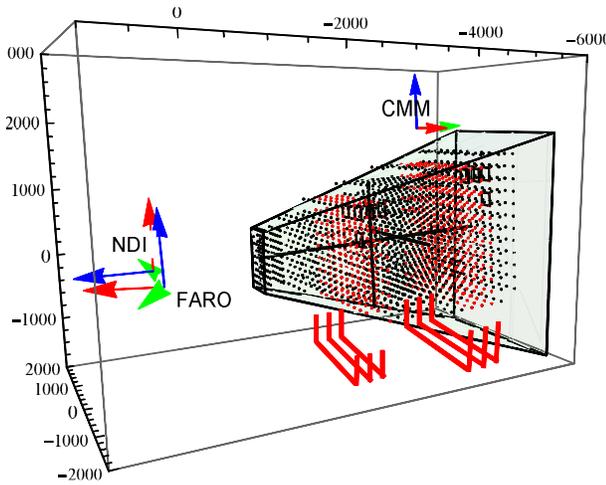


Figure 8: Measurement Volume, point grid and lines along coordinate axis ('simulated' reference target is located in marked areas)

systematic errors) of LED tracking. The real time system runs at 30 Hz and theoretically provides 90 samples for each measured position. However, the window of 3 seconds was clipped at its beginning to remove inertial effects coming from the sudden stop of the movement of the CMM. For each measured position, there remain 60 valid samples.

Noise The first experiment estimates the influence of random noise on LED tracking. The reason to estimate noise separately is the hope of lowering measurement errors by averaging in our exemplary scenario. The result is shown in Figure 9. For each position on the three densely samples lines (x-axis), an RMS residual error ϵ_{res} is computed according to

$$\epsilon_{res} = \sqrt{\frac{1}{n} \sum_{i=1}^n \|Hp_i - q_i\|^2} = \sqrt{Tr(\Sigma_{pos}^{3 \times 3})} = \sqrt{\mu^2 + \sigma^2} \quad (1)$$

over the 60 available samples (y-axis) [23]. Thereby, p_i and q_i are corresponding 3D position measurements in different coordinate frames P and Q , related by the 6DoF pose H . The root of the mean of the squared Euclidean distances $\|\cdot\|$ of the mapped point sets is equivalent to the root of the trace Tr of the 3D covariance matrix $\Sigma_{pos}^{3 \times 3}$. Assuming an isotropic error model, ϵ_{res} can also be expressed in terms of the expected deviation μ (systematic error) and its standard deviation σ (random error).

The evaluation shows that noise is constantly low in the horizontal and vertical directions. It increases however with the depth, assumably due to a decreasing amount of affected pixels on the sensors. Still, random noise is of secondary importance, with a maximum RMS of 0.06mm in the far end of the volume. More important are the systematic effects described next.

Systematic Effects The accuracy of distance measurements between LED positions on the lines is evaluated by a comparison against the corresponding distance measurements of the CMM. Noise has been eliminated before by averaging over all 60 samples per position. The result is shown in Figure 10. Three different distances are considered in the range of 10mm up to 100mm. Each of these distances was moved incrementally along the three densely sampled axes (x-axis). The deviation of two corresponding lengths is plotted at that position (y-axis). There is no systematic effect for smaller distances. The error is small and increases in magnitude for larger depths. For larger distances, however, there is a tendency to overestimate lengths in the near part and to underestimate them in

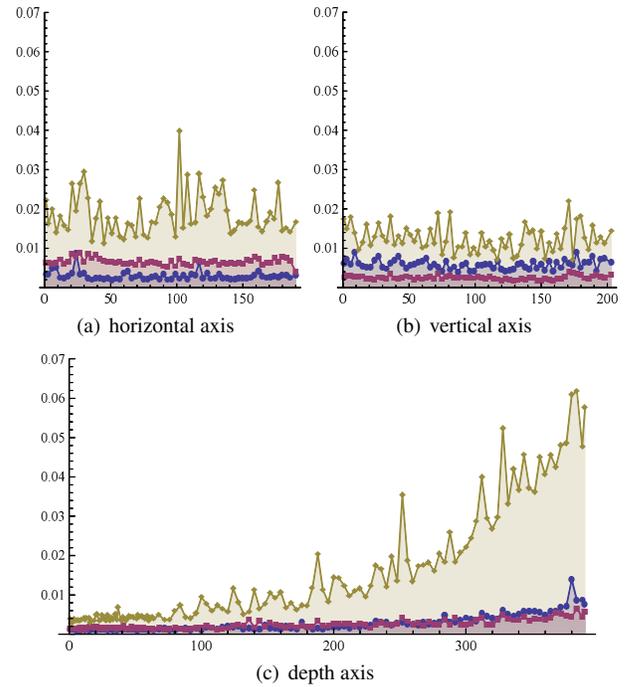


Figure 9: Noise: standard deviation [mm] of points (at 10mm distance, 60 samples per position) on coordinate axes split to \circ : horizontal, \square : vertical, \diamond : depth direction

the far part of the volume. This indicates that systematic errors take effect rather globally than locally.

Next, systematic effects shall be analyzed in the grid covering almost the entire tracking volume. By each of the three measurement systems, a point cloud as shown in Figure 8 has been recorded. Matching these point clouds by a 6DoF rigid transformation $Hp_i = q_i$ using adjustment calculation gives information about the accuracy of the individual systems [16]. Noise again has been eliminated before. The CMM actually measured its tip position, not the position of the LEDs. However, the orientation of the tip remained constant, so the desired positions of the LEDs are translated by a constant offset from the actual measurements, it will be handled implicitly by the transformation H . The same holds for the offset between the LEDs and the laser reflection target. Matching the point clouds from CMM and laser tracker results in an RMS of only 36.9 μ m. This overall RMS covers the errors of both systems. Its low value is in accordance with the specified accuracies. In conclusion, both systems, CMM and laser tracker, provide a good ground truth for testing the accuracy of the real time tracking system.

Ideally, a 6DoF rigid body transformation should suffice to match the real time tracking system with the CMM. It yields a joint RMS of 440 μ m, which is mostly caused by the real time tracker, since the error of the CMM is by an order of magnitude lower, as specified and shown above. Considering the vendor specification, Equation (1) and the fact that most grid points are located in the far part of the pyramid, this is a reasonable value. The resulting deviation vectors are shown in Figure 11, scaled by a factor of 100 for better visibility. The plot reveals that there are systematic effects, especially in the boundary areas of the volume. The computed H minimizes the RMS. Therefore, errors are lower in the center of the volume and increase towards the boundaries. Furthermore, errors on the boundaries are balanced through the minimization. According to Equation (1), the RMS value specified by the vendor and confirmed by our measurements, consists of a dominating system-

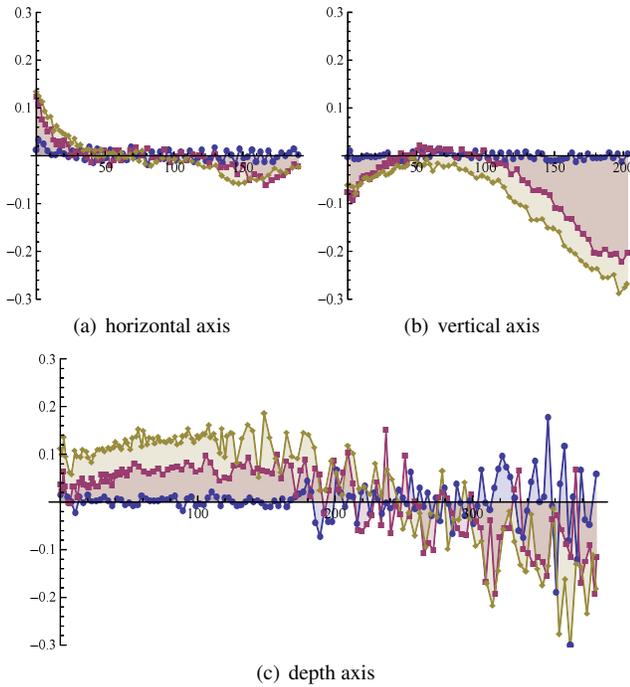


Figure 10: Systematic LED deviation [mm] from reference distance (\circ): 10, \square : 50, \diamond : 100) between two adjacent points (mean of 60 samples)

atic error μ and a minor noise component σ . Computing a more general 7DoF similarity transform (0.439mm RMS) or a 9DoF non-isotropic scaling transform (0.423mm RMS) does not decrease the error significantly. However, the 12DoF affine transform reduces the error to 0.253mm RMS, indicating that a better pre-calibration of the system on behalf of the vendor would be possible. This is similar to the system we evaluated in [14].

Noticeable, the RMS error for matching the point clouds of two LEDs (e.g. LED1, 2) is only one fourth of the global RMS (0.116mm). This indicates that there is a strong local dependency of the deviation vectors, the systematic error.

To give an intuitive example for the consequences, one might consider the probe target shown in Figure 2 which has an edge length of 5cm. The span of this target is small compared to the frequency of the systematic distortions and therefore all LEDs are affected by a similar deviation vector. This results in a direct error in the position estimate. Its orientation, however, is estimated better than the specified error model predicts because the target is less deformed. This fact also becomes obvious in Figure 10 and will have an impact in the experiment described next.

Indirect Tracking Experiment The consistency of our simulation model shall be proven by a comparison with extensive empiric measurements in the CMM setup. For this, we focus on the accuracy of indirect tracking since it has a major impact on the overall accuracy of our sample scenario and will dominate the impact of probe tracking, tip calibration, and world registration. Thus, we estimate how accurate a given POI can be estimated in world coordinates, under the influence of an error prone reference target tracking. Based on the grid of measured LED coordinates, we defined a virtual reference target consisting of four adjacent grid positions (20x20[cm]). As highlighted in Figure 8, the virtual reference target is iterated through the two marked areas in the center (near) and at the end of the pyramid (far).

Indirect tracking uses the reference target to self-localize the

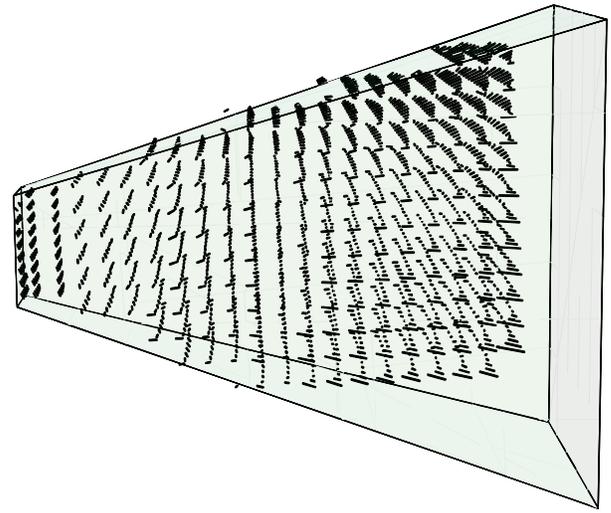


Figure 11: Systematic directed deviation vectors for each position in grid, scaled by 100

tracking system and thereby derive the estimate for the POI. This means the given POI should remain constant in world coordinates for any position of the reference target. Due to errors in LED tracking, this is not perfectly true. A 3x3 covariance can be estimated for the POI, once for the empiric LED positions, and once for the simulation, perturbing the ground truth grid positions using the vendor's uncertainty specification. In case the specification and our simulation model are correct, the resulting covariances should match.

The result is shown in Figures 12(a) for the near and 12(b) for the far location of the reference target. The plots show the confidence ellipsoids corresponding to these covariance matrices (enlarged by a factor of 100) for various POIs throughout the volume. First of all, it can be seen quite clearly that errors increase with an increasing distance of the probe with respect to the reference target, an important fact for the definition of future work processes. Furthermore, the simulated covariance matrices always enclose the empiric covariance matrices but otherwise have a similar shape. This means the simulation yields a reasonable qualitative description of the error - but in this case is an upper bound.

There is a simple explanation for this fact: the rather low frequency of the systematic distortions mentioned above. It results in a lower rotational error than the assumption of a pure Gaussian error would predict which in turn increases the performance of indirect tracking. The noise specified by the vendor is globally correct but has local dependencies as seen in the error analysis before. The overestimated rotational error in the simulation is propagated over a long axis and therefore leads to an over-estimation of the indirect tracking error.

The experiment has been repeated using other, larger sizes of the reference target, up to an edge length of 6 grid positions (120x120[cm]). Figure 13 shows the empirical and simulated minimum and maximum errors. The larger the target, the better the empiric results are approximated by the simulation.

Summary Altogether, the evaluation of the optical IR tracker in the CMM setup coincides with both, the vendor specifications and prior evaluations [23, 21]. We could verify the elementary sensor noise specification in magnitude and its increases with depth. Nevertheless, some important additional facts are revealed for the subsequent accuracy analysis in the target environment. Systematic errors dominate the overall error, especially for marginal positions. Thus, for critical applications, some outer parts of the pyramid might be clipped. Likewise, only small benefits can be expected

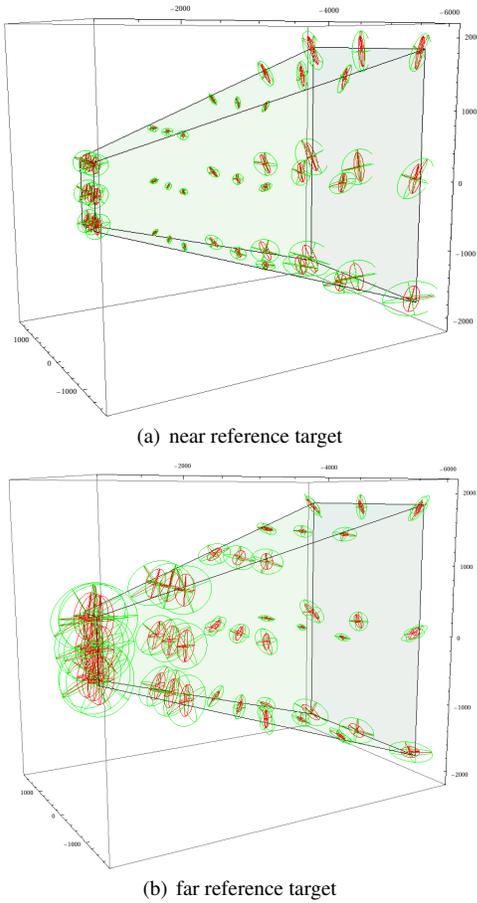


Figure 12: Comparison of simulation and empiric measurements

from computing mean values for positions measurements. Also the strong local dependency of the systematic error leads to an overestimated positional error for small targets whereas the orientational error is underestimated, compared to the assumption of a globally random error.

6.2 Verification of Use Case Scenario

For the verification of our scenario, we extend the simulation from Section 6.1 to incorporate all additional registration transformations. As explained in Section 5.1 every one of them is prone to a specific error. Using the simulation we try to derive some design decisions considering the amount of samples used for the registration procedure as well as the size of possible markers required.

The simulation results of our test case pointed out that a reference marker dimensioned to $300 \times 300 [mm]$ seems to be a good optimum to get a high accuracy and still allow handling in the aircraft.

Next we analyzed the world registration procedure using the FARO metrological system that had an RMS error of $18 \mu m$ in our test case. To understand the impact on the entire application, we use the simulation to propagate this error to the tip in world coordinates. For different numbers of samples used for the registration procedure, the resulting error in the tip is displayed in Table 1. When comparing the resulting error for 4, 6 and 8 samples it seems sufficient to rely on 4 since the over determination allows to identify user errors and the resulting error does not influence the entire system error strongly.

For the registration between the reference marker and it's CAD model we depend on the less accurate real time tracking system

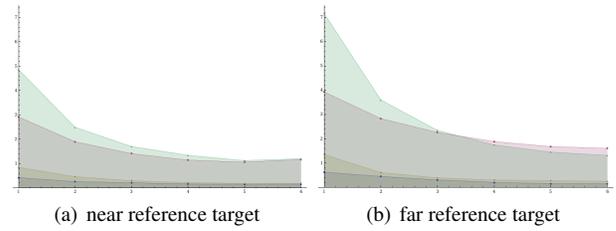


Figure 13: Comparison of POI RMS errors $[mm]$ (indirect tracking). \triangle : Simulation Max, \square : Empiric Max, \diamond : Simulation Min \circ : Empiric Min

Table 1: Propagated RMS errors $[mm]$ for the world registration

number of points	min. error	max. error
3	0.0790	0.690
4	0.0590	0.488
6	0.0547	0.475
8	0.0474	0.405

with an RMS point error of $200 \mu m$, computed from the vendor specification using (1) $(\sqrt{0.10^2 + 0.10^2 + 0.15^2})$. To understand the impact on the entire application, we again use the simulation to propagate this error to the tip in world coordinates. In contrary to the prior analysis and due to the reduced accuracy we decided to rely on more samples and to compare different designs of the probing device. Table 2 indicates that for the registration of the reference target the type of probe is less important.

Table 2: Propagated RMS errors $[mm]$ for the reference target

used probing device	min. error	max. error
small probe ($5 \times 5 [cm]$)	0.208	2.23
large probe ($10 \times 10 \times 5 [cm]$)	0.192	2.11

A typical setup we want to validate has the reference target in the far field (Figure 8). Following the results from Figure 12, we prohibit the use of the probe too far away from the reference target and therefore clip the front part of the pyramid to avoid occlusion and a too high error value.

As Table 3 illustrates, the error in the worst case estimation of our scenario ranges between $0.79 mm$ and $4.86 mm$. As already stated, the error increases with the distance between probing device and reference target, as in Figure 12(b). This becomes clear when we use four small 3LED reference targets distributed in the working volume (multi target) - the maximum error drops to $1.57 mm$ whereas the minimum error is nearly unchanged. This coincides with the basic rule in metrology to always measure inside the cloud of reference points used for registration [16]. Also we simulated the expected error for an alternative version of the hardware. This was specified with less than a third of the 3D position error¹¹.

6.3 Validation of Use Case Scenario

As proposed we validate our simulation results in the real environment. Starting with the noise of the direct and indirect probe tip tracking. In the direct probe tracking the results matched the expected values and the standard deviation for various measurements was $0.03 - 0.04 [mm]$.

When measuring the tactile points of one of the reference targets, the back projection error was too large. Recalibration with the 25

¹¹NDI Optotrak Pro System, RMS error $[mm]$ for horizontal / vertical / depth directions (depending on depth): $0.02/0.02/0.06 [mm]$ (at $2m$), $0.06/0.06/0.15 [mm]$ (at $6m$)

Table 3: Simulated RMS errors [mm] for the tip position

simulation setup	min. error	max. error
far reference target	0.789	4.86
far reference target (pro)	0.154	1.43
far multi target	0.710	1.57
far multi target (pro)	0.137	0.377

tactile points solved that issue. Also for the just calibrated reference targets, the accuracy was as expected, the error ranged from 1mm to 3mm with a standard deviation of approximately 0.5mm over 100 samples - depending on the distance between reference target and probe. If the application requires a precision below 1mm, the more expensive “pro” system needs to be used. Also, the degradation in the quality of the calibration of the reference targets over time will have to be validated in a more detailed system analysis.

7 DISCUSSION AND FUTURE WORK

We demonstrated that using a simulation framework, design decisions in the definition phase of an industrial tracking system become more transparent. Since complex tracking scenarios cannot be covered by elementary uncertainty specifications, it is a great help to understand the propagation of errors throughout the complete chain of errors. We developed an integrated approach to cover the verification and validation of the system, based on the provision of an elementary uncertainty specification. Besides having a detailed insight into the error behavior of the system, it is also possible to define run-time checks that allow to distinguish between correct system behavior and various error cases. Since we proved our simulation approach to yield realistic results in the exemplary setup, similar applications can be verified and validated with less testing effort in the future. Additionally, various different hardware platforms and concepts can be easily benchmarked even without the need to use real hardware.

In the future we are planning to apply our new concept to other industrial tracking scenarios for which measurement data is already available [14]. Doing so will substantiate the feasibility of our generalized approach. Additional complexity arises when looking at heterogeneous systems. In our use case a single real time tracking system is used and therefore there is no need to consider time. Combining measurements with different timestamps requires inter-/extrapolation and therefore the error simulation would have to be extended, too. Finally, a graphical tool to configure and examine the test cases and the simulation results would be desirable. It should be possible to define the poses for tracking devices, markers and other entities of the SRG directly in a 3D environment.

ACKNOWLEDGEMENTS

This work was partially supported by the German Federal Ministry of Education and Research (AVILUS project, grant no. 01 IM 08 001 A) as well as the German Federal Ministry of Economics and Technology (AiF-FV 14756).

REFERENCES

- [1] B. D. Allen and G. Welch. A general method for comparing the expected performance of tracking and motion capture systems. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST'05)*, pages 201–210, New York, USA, 2005.
- [2] ASME. Guide for verification and validation in computational solid mechanics. Technical report, The American Society of, October 2006.
- [3] M. Bauer, M. Schlegel, D. Pustka, N. Navab, and G. Klinker. Predicting and estimating the accuracy of vision-based optical tracking systems. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Santa Barbara, USA, October 2006.
- [4] K. Daniilidis. Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research*, 18(3):286, 1999.

- [5] J. M. Fitzpatrick, J. B. West, and C. R. M. Jr. Predicting error in rigid-body, point-based registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702, 1998.
- [6] M. Haller, M. Billinghurst, and B. Thomas. *Emerging technologies of augmented reality: interfaces and design*. Idea Group Pub., 2007.
- [7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.
- [8] H. Hastedt. Monte-carlo-simulation in close-range photogrammetry. In *ISPRS Symposium Commission V, Istanbul, Turkey, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, volume 35, pages 18–23, 2004.
- [9] W. Hoff and T. Vincent. Analysis of head pose accuracy in augmented reality. In *IEEE Transactions on Visualization and Computer Graphics*, volume 6(4), pages 319–334. IEEE Computer Society, 2000.
- [10] R. Holloway. *Registration Errors in Augmented Reality Systems*. PhD thesis, University of North Carolina, 1995.
- [11] B. Horn, H. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4):629–642, 1987.
- [12] ISO/IEC. Guide 98-3:2008: Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement (GUM). Published, International Organization for Standardization (ISO), Geneva, Switzerland, September 2008.
- [13] P. Keitler, D. Pustka, M. Huber, F. Ehtler, and G. Klinker. *Engineering of Mixed Reality Systems*, chapter Management of Tracking for Mixed and Augmented Reality Systems. Dubois, E. and Gray, P. and Nigay, L., 2009.
- [14] P. Keitler, M. Schlegel, and G. Klinker. Indirect tracking to reduce occlusion problems. In *Advances in Visual Computing, Fourth International Symposium, ISVC 2008 Las Vegas, USA, December 1-3*, volume 5359(2) of *Lecture Notes in Computer Science*, pages 224–235, Berlin, 2008. Springer.
- [15] S. Lieberknecht, S. Benhimane, P. Meier, and N. Navab. A dataset and evaluation methodology for template-based tracking algorithms. In *ISMAR*, 2009.
- [16] W. Niemeier. *Ausgleichsrechnung - Statistische Auswertungsmethoden*. Walter de Gruyter Verlag, Berlin, Germany, second edition, 2008.
- [17] K. Pentenrieder, P. Meier, G. Klinker, and M. Gmbh. Analysis of tracking accuracy for single-camera square-marker-based tracking. In *Proc. Dritter Workshop Virtuelle und Erweiterte Realitt der GI-Fachgruppe VR/AR, Koblenz, Germany*, 2006.
- [18] D. Pustka, M. Huber, M. Bauer, and G. Klinker. Spatial relationship patterns: Elements of reusable tracking and calibration systems. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR'06)*, October 2006.
- [19] R. Rohling, P. Munger, J. M. Hollerbach, and T. Peters. Comparison of relative accuracy between a mechanical and an optical position tracker for image-guided neurosurgery. In *Journal of Image Guided Surgery*, pages 277–282. Wiley and Sons Inc, 1994.
- [20] K. Satoh, K. Takemoto, S. Uchiyama, and H. Yamamoto. A registration evaluation system using an industrial robot. In *IEEE/ACM International Symposium on Mixed and Augmented Reality, 2006. ISMAR 2006*, pages 79–87, 2006.
- [21] J. Schmidt, D. Berg, H. Ploeg, and L. Ploeg. Precision, repeatability and accuracy of Optotrak optical motion tracking systems. *International Journal of Experimental and Computational Biomechanics*, 1(1):114–127, 2009.
- [22] Optical 3d measuring system - imaging systems with point-by-point probing. VDI/VDE guideline 2634/1, 2002.
- [23] A. Wiles, D. Frantz, D. Swart, et al. Ndi accuracy assessment kit guidelines. Technical report, Northern Digital Inc., 2005.