Extended-Range Hybrid Tracker and Applications to Motion and Camera Tracking in Manufacturing Systems
Dan Zetu, Pat Banerjee, and Darren Thompson

Abstract—Extended- or long-range tracking effectiveness is crucial for the automation of manufacturing systems. In this paper, we conceptualize and develop a prototype long-range hybrid tracker based on a combination of a laser tracker and a magnetic tracker and apply the concept to the following two applications: 1) extended-range human motion tracking on factory floors and 2) factory floor object reconstruction from camera images. The easily portable system not only utilizes the strengths of a laser tracker in tracking mobile objects over long ranges in large environments, such as a manufacturing shop floor and the strength of a magnetic tracker to compensate for violation of line-of-sight constraint, but it also reduces the overall cost by reducing the number of expensive beacons required by the laser tracker. The hybrid tracker assists in the development of two concepts: 1) real-time synchronization of human head and hand motion in a manufacturing environment with those of an avatar in a virtual manufacturing environment and 2) a mathematically simpler and practical camera self-calibration technique for the creation of three-dimensional objects in a virtual environment from camera images.

Index Terms—Hybrid tracker, laser tracking, magnetic tracking, stereo reconstruction, virtual reality applications.

I. INTRODUCTION

AUTONOMOUS navigation of mobile robots and material handling equipment (such as automated guided vehicles or forklifts) is often a prerequisite for automation of manufacturing systems. In order to achieve autonomous navigation of such components of manufacturing systems, they need to know at any time where they are located within the environment, with respect to a global coordinate system [20]. Similarly, in virtual reality (VR)-aided manufacturing systems design and maintenance applications, it is necessary to capture the motion of human participants in order to replicate it on avatars within virtual environments (VE’s) representing specific manufacturing systems. This motion has to be often captured over a longer range than the ranges of current tracking systems for VR applications. A survey of the existing position trackers used in VR can be found in [12] and [25].

Currently, extended-range trackers are employed for tracking mobile robots [7], [23], [13]. The most widely used long-range tracking systems in robotics are active beacon systems. The biggest disadvantage of active beacon navigation systems is the line-of-sight constraint (LOS) [7]. There may be instances when tracking a certain part (such as the end-effector) of a robot is necessary, and this part cannot be visible to the tracking system, unless multiple beacons are placed within the motion environment, thus increasing substantially the cost of the tracking system.

In VR applications, the task of a tracker is to report the position and orientation of a user’s head and hand. Accordingly, the VR system updates the perspective display to make it consistent with the user’s viewpoint. There are multiple types of tracking systems used in VR: magnetic, optical, mechanical, acoustic, and inertial. Each of these trackers has advantages and disadvantages. For example, magnetic trackers have no LOS constraints, but their accuracy decreases dramatically with increase in distance from transmitter and is also influenced by metallic objects in the neighborhood. Optical trackers are very fast and accurate, and they are also immune to magnetic interference, but their use is restricted by the LOS constraint. Mechanical trackers, based on linkages, are very accurate, but their work is severely restricted within a small-range volume (determined by the geometry of linkages). Acoustic (ultrasonic) trackers are relatively cheap and accurate, but they also have limited range and LOS restriction.

The above considerations suggest combining some of the advantages offered by individual tracking systems to design a hybrid tracker for autonomous navigation in real manufacturing environments and human motion in VE’s. In this paper, a hybrid tracker, based on a combination between a laser tracker and a magnetic tracker, is described. The laser tracker used is an active beacon system [36], and the magnetic tracker employed is called MotionStar® [37]. The laser tracker has the advantage of enabling accurate tracking of position and orientation over long ranges (the system we use has a maximum range of 100 m, but through serialization of multiple such systems, unlimited range can be obtained). The magnetic tracker enables tracking of multiple parts of an object without problems due to occlusion. Moreover, when the LOS constraint for the laser tracker is temporarily violated, the magnetic tracker can compensate for it. This is an important advantage since it reduces the number of beacons or landmarks that have to be placed within the environment, thus reducing substantially the cost of tracking systems.

This paper is organized as follows. Related work in hybrid tracking is discussed in Section II. Section III describes the geometry of the hybrid (laser and magnetic) tracking system.

1CONAC by MTI Research, Westford, MA 01886 USA.
2Manufactured by Ascension Technologies Corporation, Burlington, VT 05402 USA.
Section IV describes the application of the proposed hybrid tracker to motion tracking during real-time synchronization of real environments and VE’s. Section V covers three-dimensional (3-D) object reconstruction from camera images and describes our camera self-calibration procedure using extended-range tracking. Section VI is devoted to conclusions and future directions.

II. RELATED WORK

Due to their contribution to the end-to-end latency of VR systems, tracking systems have received a great deal of attention among VR research community. Hybrid tracking has been explored mostly in the area of augmented reality (AR), where accurate registration between real environment and virtual objects superimposed on it is critical. In [2], the need for hybrid tracking in AR is stressed, especially for outdoors applications. Most of the tracking in AR is being performed with the aid of video cameras, by tracking fiducial marks (placed at known locations in the environment) using computer vision techniques. Despite the accuracy of these techniques, they are slow (due to the necessity of searching the marks by scanning the image pixel by pixel), their range is limited by the placement of the fiducial marks and are not robust to occlusions of the fiducial marks. In [26], a hybrid tracker is described, which combines computer vision-based tracking with inertial tracking. Since it is well known that inertial trackers exhibit drift with time (their errors increase over time [3]), their output is corrected by using vision-based tracking. Another hybrid system has been proposed in [17], where it has been proven that by combining two types of vision-based tracking, called “inside-out” (camera(s) mounted on the head of the user and fiducials mounted at known locations in the environment) and “outside-in” (cameras mounted at known locations in the environment and fiducials mounted on the user’s head), the uncertainty in head pose (position and orientation) estimation is considerably decreased. In [28], the accuracy of a magnetic tracker is improved by augmenting it with a passive image-based system that observes known fiduciary marks in the real world. At the same time, the magnetic tracker measurements help in reducing the search area of the fiducials in two-dimensional (2-D) images captured by head-mounted cameras, thus reducing the latency of the hybrid tracker. Other hybrid systems have been previously proposed in [5], [9], and [14]. For example, in [5] and [9], combinations between inertial and optical technologies are described in terms of accuracy and end-to-end latency. In [14], an inertial system is aided by angular position sensors. None of these applications address the problem of tracking motion in large environments, such as a factory floor. For this kind of application, active beacon systems are very suitable, due to their accuracy and extended range but, due to the LOS constraint, usually a large number of beacons has to be mounted on the factory floor. We overcome this disadvantage by using a magnetic tracker in combination with a laser tracker.

III. DESCRIPTION OF THE HYBRID TRACKING SYSTEM AND GENERIC METHODOLOGY FOR MOTION TRACKING

As mentioned in Sections I and II, our hybrid tracker for motion tracking is a combination of a laser tracker and a magnetic one. The laser tracker provides high accuracy and update rate for high ranges (0–100 m), but its use is restricted by the LOS constraint. On the other hand, the magnetic tracker does not require LOS, but it is accurate only within small working volumes. The laser tracker is based on triangulation of laser signals emitted by two beacons and received by one or more position transponders (PT’s), attached to the moving object. One PT can report only the position with respect to one beacon so, in order to retrieve the orientation, one has to employ three PT’s rigidly mounted on a special fixture. The advantage of using a magnetic tracker is its suitability to applications with frequent occlusions between transmitter and receiver. The magnetic tracker employed in our tracker uses pulsed direct current (dc) magnetic fields instead of alternate current (ac) magnetic fields (which are being used by Polhemus, Inc. magnetic trackers and older versions of Ascension Technologies trackers). DC fields are significantly less susceptible to metallic distortion than ac fields. However, dc-based magnetic trackers are susceptible to interference with magnetic fields generated by ferromagnetic objects (such as computer monitors or dc motors, see [27]). Even though it is hard to estimate up front the probability of encountering such objects during a motion sequence, it is reasonable to assume that in most cases the wearer of a magnetic tracker will not be in the immediate proximity of ferromagnetic objects that would catastrophically affect the tracker’s output. Overall, by weighing its advantages, the dc-based magnetic tracker remains a reliable magnetic tracker for motion capture in manufacturing environments. By using a Kalman filter [19] to minimize the external effects on its performance, reasonable results can be obtained, as will be seen later in this paper. The advantages of incorporating a magnetic tracker into our hybrid tracker are as follows.

It can track multiple targets without worrying about occlusions between transmitter and receivers. Since its behavior is not influenced by an LOS constraint, the magnetic tracker can be used as a backup, when the LOS between laser tracker’s beacons and PT’s is temporarily occluded. This enables reduction in the number of beacons (landmarks) used for motion tracking, thus reducing the cost of the tracking system. Details are provided below.

Typically, magnetic tracker’s receivers are placed on components whose motion trajectories have to be captured and cannot be “seen” all the time by beacons of the laser tracker. The PT’s of the laser tracker are mounted on the moving objects, in a location that is always visible to the transmitting beacons. The positions of the tracked components (i.e., the components equipped with a magnetic receiver) are reported either with respect to a beacon’s coordinate system or to a coordinate system attached to the motion environment, termed world coordinate system (WCS) (in this case, the transformation between the WCS and beacon coordinate systems is known a priori). For this purpose, one of the magnetic receivers is rigidly attached to the PT’s, so the transformation between this receiver and PT’s is invariant as the tracked object moves. The magnetic transmitter is also placed on the moving object.

The described hybrid tracker has the advantage of being easily portable, unlike other trackers currently in use, such as UNC HiBall [31], which requires a large number of beacons (LED’s) mounted on the ceiling and whose range is limited.
by the number of such beacons. The tracker described in [31] has the advantage that, by mounting a large number of closely located LED’s on the ceiling, one will have less problems with LOS but, in order to increase the tracking range, the cost of the system rises significantly, and the system becomes less portable. By being equipped with only two stationary beacons (at least for now), our tracking system is expected to be more susceptible to LOS problems. By mounting the beacons in optimal locations (to minimize the likelihood of violating the LOS constraint) and by using also the magnetic tracker to aid the laser tracker temporarily (when the LOS constraint is violated), we expect the impact of LOS problems to be minimal. The following two questions arise. 1) How accurate are the tracker outputs when the magnetic tracker aids in circumventing LOS problems? 2) For how long can the violation of the LOS constraint be tolerated so that the position estimates fall within acceptable accuracy limits? These questions will be addressed through an example in Section IV.

The generic geometry of our hybrid tracker is depicted in Fig. 1. In this figure, the case when the positions of the tracked components are reported with respect to a WCS is illustrated. Note that in Fig. 1, only one beacon (B) of the laser tracking system is shown. In reality, the laser tracker has two beacons, but the position is reported with respect to a coordinate system associated with one of the beacons, so in order to simplify the figure, only this beacon is shown. The notations employed in Fig. 1 are summarized in Table I.

In Fig. 1, only one tracked component (denoted Ri) is shown, for the purpose of clarity. Our hybrid tracker can track as many components as the magnetic tracker allows (up to 40 targets). The position of the tracked component Ri, w.r.t. WCS, is represented by the vector \( \mathbf{x} \), and w.r.t. the magnetic transmitter is given by \( \mathbf{x}_{\text{m}} = (x_m, y_m, z_m)^T \). As can be seen from Fig. 1, \( \mathbf{x} \) and \( \mathbf{x}_{\text{m}} \) can be related by the following equation:

\[
\mathbf{x} = \mathbf{T}_{\text{MT}}^{-1} \mathbf{T}_{\text{TT}} \mathbf{T}_{\text{BT}} \mathbf{T}_{\text{WB}} \mathbf{x}_{\text{m}}. \tag{1}
\]

The vector \( \mathbf{x}_{\text{m}} \) is measured by the receiver Ri w.r.t. the magnetic transmitter. So (1) is the basic equation for tracking a component within WCS. For tracking the object globally (as a whole), only PT is used, therefore the magnetic tracker is not needed (unless the LOS constraint is violated). The hybrid tracker described here can track an unlimited number of moving objects. For each object, a distinct set of PT’s and a separate magnetic tracker is needed. The examples provided in this paper consider only a single tracked object, without loss of any generality.

### A. Hybrid Tracker Precalibration

In order to compute the position of a tracked component with respect to WCS, one needs the transformation between PT coordinate system and the coordinate system associated with the magnetic receiver that is rigidly attached to the PT’s (labeled RT in Fig. 1). This transformation is labeled \( \mathbf{T}_{\text{TT}} \) in Fig. 1 and is invariant as the PT-RT ensemble moves. In order to compute \( \mathbf{T}_{\text{TT}} \), precalibration of the hybrid tracker is performed before starting the motion tracking process. The geometry associated with precalibration is depicted in Fig. 2.

The notations used in Fig. 2 are summarized in Table II, for a generic case, as well as for two applications that demonstrate the use of our hybrid tracker (human motion capture and camera tracking for object reconstruction—applications described in Sections IV and V, respectively). In Fig. 2, the coordinate transformations and coordinate systems specific only to (or at least closely related to) camera tracking are written with a different font and are represented by dashed lines.

Hybrid tracker precalibration is performed as follows. The tracked object is placed in two arbitrary locations within the environment (care must be taken so that no ferromagnetic objects are located in the neighborhood to ensure that magnetic readings are not distorted), from where readings from PT’s and RT are collected with the object stationary. The two consecutive positions are denoted by indices \( a \) and \( b \) in Fig. 2. The relative transformations between positions \( a \) and \( b \) can be written as in (2) and (3) below (for both PT and RT).

Let

\[
\mathbf{T}_{\text{Pab}} = \mathbf{T}_{\text{BTB}} \mathbf{T}_{\text{BTB}}^{-1}. \tag{2}
\]
and let

\[ T_{\text{Rab}} = T_{\text{MTb}} T_{\text{MTDa}}^{-1}. \] (3)

From Fig. 2, the following equation can be written:

\[ T_{\text{Rab}} = T_{\text{TT}} T_{\text{Pab}} T_{\text{TT}}^{-1}. \] (4)

Equation (4) follows from the fact that the transformations involved form a closed loop. From (4), it follows that

\[ T_{\text{Rab}} T_{\text{TT}} = T_{\text{TT}} T_{\text{Pab}} \] (5)

from which \( T_{\text{TT}} \) is computed. Equation (5) is an equation of the form \( AX = XB \), typically encountered in hand-eye calibration in robotics applications. To solve (5), we use the method proposed in [29].

**B. Violation of the LOS Constraint**

In order to track all components with respect to WCS, the transformation between beacon (B) and PT coordinate system has to be known and is given by the laser tracker. In order to recover this transformation, all three PT’s have to be visible at any time by both beacons of the laser tracker. In tracking mobile robots on the factory floor by using active beacon systems, usually beacons are placed at optimal locations throughout the environment [7]. This can be easily done when the paths are predefined or are expected to take place in well-known areas, but also increases the cost of tracking systems.

When a tracker is used to capture unpredictable motion (such as human motion), one cannot design a priori an optimal configuration of beacons to prevent violation of the LOS constraint. To get around this problem, we can use the magnetic tracker (specifically the receiver attached to PT’s - RT in Fig. 1) to back-up the system when the LOS of the laser tracker is temporarily occluded. Consider again Fig. 2. Let us assume that the tracked component moves from position \( a \) to position \( b \). In position \( a \), all three PT’s are visible, and therefore the transformation \( T_{\text{BTa}} \) is correctly reported. In position \( b \), at least one PT is occluded. In this case, the transformation \( T_{\text{BTb}} \) can be recovered from the previous estimate of the PT’s position and orientation (\( T_{\text{BTa}} \)) and the relative motion undertaken by magnetic receiver RT (denoted as \( T_{\text{Rab}} \)), by the following equation:

\[ T_{\text{BTb}} = T_{\text{TT}}^{-1} T_{\text{Rab}} T_{\text{TT}} T_{\text{BTa}}. \] (6)

In (6), \( T_{\text{Rab}} \) is measured w.r.t. the magnetic transmitter. Equation (6) is valid when the magnetic transmitter remains fixed relative to WCS or its motion w.r.t. WCS is negligible by comparison of RT motion w.r.t. WCS. For example, when tracking human motion, the magnetic transmitter is placed on the back of the human and RT on the user’s head. When the human operator bends (and thus PT’s are not visible from the beacons), the magnetic transmitter remains relatively fixed. The potential violation of this assumption is considered while designing the Kalman filter that deals with LOS constraint violations (described in Section IV), by scaling up the measurement noise uncertainty.

**C. Operating the Hybrid Tracker**

When retrieving position and orientation information by fusing data provided by two or more sensors, typically the assumption that measurements are available simultaneously from all sensors is made. In reality, this is almost never the case, due to different update rates of the various sensors. In our case, measurements from the laser and magnetic trackers are fed to a 300-MHz Pentium PC via serial cables and from there to an SGI workstation that performs all the calculations for position and orientation estimates. Since tracking is initiated only when the first data packet arrives from both sensors, communication overhead is not relevant for the time increment between two consecutive measurements. The interval between two measurements of the laser tracker is 22 ms. When using a single receiver, the update rate of the magnetic tracker is 5 [ms] and increases by the same amount as a new receiver is added. The temporal diagram shown in Fig. 3 depicts the succession of measurement packets as those arrive to the SGI workstation when using three receivers of the magnetic tracker (in this case the update rate is 15 [ms]).

As can be seen from Fig. 3, the measurements arriving from the laser and the magnetic trackers are not synchronous. This introduces an error in estimating the true position of a tracked
TABLE II
EXPLANATION OF NOTATIONS IN Fig. 2

<table>
<thead>
<tr>
<th>Notation</th>
<th>Generic</th>
<th>Human motion</th>
<th>Camera tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT&lt;sub&gt;a,b&lt;/sub&gt;</td>
<td>PT coordinate system in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;, respectively</td>
<td>Same (PTs are head-mounted)</td>
<td>Same</td>
</tr>
<tr>
<td>T&lt;sub&gt;RTa,b&lt;/sub&gt;</td>
<td>Coordinate transformation matrix between beacon coordinate system (B) and PT coordinate system in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;, respectively</td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>T&lt;sub&gt;rab&lt;/sub&gt;</td>
<td>Transformation undertaken by PTs from position &lt;i&gt;a&lt;/i&gt; to position &lt;i&gt;b&lt;/i&gt; (w.r.t. B)</td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td>T&lt;sub&gt;rab&lt;/sub&gt;</td>
<td>Transformation undertaken by receiver RT from position &lt;i&gt;a&lt;/i&gt; to position &lt;i&gt;b&lt;/i&gt; (w.r.t. the magnetic transmitter)</td>
<td>Same</td>
<td>Transformation undertaken by camera from position &lt;i&gt;a&lt;/i&gt; to position &lt;i&gt;b&lt;/i&gt;</td>
</tr>
<tr>
<td>&lt;sup&gt;WCS/&lt;/sup&gt;T&lt;sub&gt;WB&lt;/sub&gt;</td>
<td>World Coordinate System/ Transformation between WCS and beacon (B) – used for the purpose of registering real environments with virtual environments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCS</td>
<td>Camera Coordinate System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T&lt;sub&gt;FF&lt;/sub&gt;</td>
<td>Invariant transformation between PT and RT (parameter to be calibrated)</td>
<td>Same</td>
<td>Invariant transformation between PT and CCS</td>
</tr>
<tr>
<td>CCS&lt;sub&gt;a,b&lt;/sub&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>CCS in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;</td>
</tr>
<tr>
<td>T&lt;sub&gt;Wca,b&lt;/sub&gt;</td>
<td>N/A</td>
<td>N/A</td>
<td>Transformation between WCS and CCS&lt;sub&gt;a,b&lt;/sub&gt;</td>
</tr>
<tr>
<td>RT&lt;sub&gt;a,b&lt;/sub&gt;</td>
<td>RT in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;, respectively</td>
<td>Head-mounted magnetic receiver in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;</td>
<td>RT in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;, respectively</td>
</tr>
<tr>
<td>T&lt;sub&gt;MTa,b&lt;/sub&gt;</td>
<td>Coordinate transformation matrix between the magnetic transmitter and RT in positions labeled &lt;i&gt;a&lt;/i&gt; and &lt;i&gt;b&lt;/i&gt;, respectively</td>
<td>Same</td>
<td>Same</td>
</tr>
</tbody>
</table>

*Note: In the case of camera tracking, RT<sub>a,b</sub> and CCS<sub>a,b</sub> are distinct coordinate systems. For reasons of compactness, this fact is omitted in fig. 2.*

![Temporal diagram of the hybrid tracker measurements.](image)

Component, since it is not possible to collect a measurement from both trackers at exactly the same moment in time. The fact that there is no constant offset between readings complicates the problem. Due to the small temporal difference between measurements collected from the two sensors and due to the fact that the expected number of magnetic sensors typically used in our applications is between 2–6, the errors are not expected to be significant in comparison to the errors inflicted by the noise in the measurements. Consider the case shown in Fig. 3. In the current stage of our hybrid tracker, if in between two successive readings from the laser tracker there is only one reading from the magnetic tracker, this one is considered in the calculations. If two or more readings appear, these are first averaged to obtain a more realistic estimate.

The position of a tracked component is estimated through a Kalman filter [19]. When using Kalman filters in tracking applications, the following steps have to be performed before process initiation [8], [24]:

- identification of state variables and measurement parameters;
• choice of a dynamic model (the dynamic model depends on the particular application and on the nature of motion being captured);
• model the process and measurement noise;
• initialize state variables and error covariance.

The steps mentioned above are the same for time-varying systems (measurements are collected when the tracked object is in motion) and time-invariant systems (measurements are collected when the tracked object is stationary). Our hybrid tracker can be used for both types of systems.

In our case, two Kalman filters are alternately used, depending on whether or not the LOS constraint of the laser tracker is violated. When there is occlusion between the beacon and any PT, the laser tracker stops sending data to the interface module (each PT has its own interface module in order to increase the update rate). This case is tested by monitoring the time interval $\Delta t_s$ elapsed from previous measurement. If $\Delta t_s$ exceeds 26 [ms], the LOS constraint is considered violated (recall that the update rate of the laser tracker is 22 [ms] and we allow 4 [ms] for possible communication glitches). In this case, the system switches to the alternate Kalman filter that uses the same state and measurement models, but has a larger initial error covariance and different measurement noise model due to the fact that the accuracy of the magnetic receiver RT (that backs up the laser tracker) is expected to be lower than the one of the laser tracker. When the LOS between beacons and PT’s is free, the laser tracker starts outputting measurements automatically and a switch to the regular Kalman filter is performed. Filtering is resumed with the predicted state variables and error covariance given by the back-up filter, instead of the same values before occlusion of the LOS (we found out that this approach is more appropriate because the motion estimation is smoother and the amount of jitter is reduced). The only change is that the direct laser tracker measurement is used instead of (6). The generic methodology of operating the hybrid tracker when capturing motion can be summarized as in the diagram shown in Fig. 4.

IV. APPLICATION TO HUMAN MOTION

Human motion is an example of using the hybrid tracker with time-varying systems. Capturing human motion in manufacturing environments in order to be replicated in VE’s is a challenging task. In VR applications, typically head and hand of a user are tracked in order to update the perspective. In order to achieve realistic human motion in VE’s, more components of a human body have to be tracked (such as torso and joints). In order to illustrate the application of our hybrid tracker to human motion capture, we limit ourselves to tracking only the head and hand of a human on a factory floor, replicated by an avatar in a VE representing the real factory floor. VE is a priori registered with the real environment.

The three PT’s of the laser tracker and one receiver of the magnetic tracker (RT in Fig. 1) are rigidly mounted on a fixture with the shape of a hat, mounted on the user’s head. The magnetic transmitter is placed in a backpack, located on the back of the user or, when motion takes places within a small volume (but at a large distance from a reference point, thus requiring laser tracking as well), it can be placed in a fixed position, close to the human operator.

The user’s hand is tracked by means of a magnetic receiver, attached to the wrist, based on which the hand position with respect to WCS can be computed. The laser tracker gives the head position.

The measurements performed by the tracker over time are noisy. It is reasonable to assume [8], [32], [4] that the process of position estimation is driven by normally distributed noise. In order to optimally estimate the coordinates of hand position with respect to WCS in a noisy environment, we design a Kalman filter, as mentioned in Section III. The precise dynamic model of hand and head motion is unknown, but the Kalman filter can provide very good results even for this kind of application [32]. Also, the Kalman filter has been found ([4]) to perform well even when the assumptions of normal distribution of noise representing the uncertainties in the measurements in the model are violated.
The use of a Kalman filter for motion capture requires a motion model. Unfortunately, it is almost impossible to obtain an accurate model for the hand and/or head motion. To get around this problem, researchers have used different models to approximate head and/or hand motion. In [32], the position-velocity model (defined in [8]) has been used, which assumes motion takes place at constant velocity and models acceleration as white noise. In [21], it is assumed that head rotations are infrequent and that angular speed and angular acceleration are nonzero only during infrequent change in viewing direction. These assumptions led to the choice of an integrated Gauss–Markov process to model the head movement. In [15], a hand motion model with constant acceleration has been used. All these approximations provide satisfactory results, with occasional overshoot when sudden change of direction or velocity occurs. We have used the acceleration model [15], [6] to approximate hand and head motion. The Kalman filter for hand and head tracking is briefly described in Appendix A.

In order to determine the performance of the hybrid tracker, the magnetic receiver that records the hand position is positioned initially at some known world locations, in order to determine the process and measurement noise covariance matrices and respectively. Measurements are collected and the filter is run offline. The error is the difference between the estimated and actual positions of the tracker. A cost function is defined as the sum of the squared errors at each time step. Through the minimization of the cost function, matrices and are computed. The Kalman filter error covariance matrix is assumed to be diagonal. The diagonal elements of are initialized to some large values (2 for the elements corresponding to position, 50 for velocity, and 60 for acceleration). The elements corresponding to velocity and acceleration are initialized to higher values than the ones corresponding to position because motion tracking starts with the user being stationary.

When tracking human motion, one does not have available ground truth data, since it is impossible to predict exactly the movement path. To assess the accuracy and consistency of our Kalman filter model, we have captured motion sequences with frequent changes of direction and monitor the difference between Kalman filter predictions and real tracker measurements. Metallic objects were present close to the movement path to illustrate a relatively insignificant impact on our model. The results are shown in Figs. 5 and 6.

Figs. 5 and 6 show that there is a reasonable consistency between Kalman filter predictions and real tracker measurements, with slight overshoot or undershoot when a sudden change of direction occurs. The type of motion depicted in Figs. 5 and 6 takes place in very unfavorable conditions for
our motion model. Typically, we do not expect such frequent changes of direction, and therefore the overshoot or undershoot will be reduced. We performed multiple experiments and all provided similar results. The maximum overshoot encountered was about 7 cm. Fig. 6 depicts a special case, when the laser tracker is occluded for approximately 9 s. This occlusion happens to coincide with a change in direction of motion. When such situations occur, the measurement noise level is scaled up to reflect the additional uncertainty in position estimation when using the relative motion of the magnetic tracker (6) instead of the laser tracker (11). As can be seen, even though it is still at an acceptable level, the amount of overshoot in this case is larger than the typical overshoot when changing direction of motion, as shown in Fig. 5.

The results presented in this section show the consistency of our Kalman filter and show that the motion model we have chosen gives sufficiently accurate results even in unfavorable cases encountered in human motion. Also, as was shown in Fig. 6, the magnetic tracker can effectively back-up the laser tracker when it is occluded for an amount of time (in our case, 9 s). We do not recommend using the magnetic tracker to compensate for the violation of LOS constraint for more than 10 s (which is sufficient in most cases). The presence of metallic objects definitely affects the accuracy of the tracker, but not to an unacceptable extent. Based on the results obtained, we conclude that our hybrid tracker can be effectively used for extended-range motion capture in manufacturing environments, for applications that can accept a maximum error of 5–7 cm in position estimation.

V. APPLICATION TO 3-D FACTORY OBJECT RECONSTRUCTION FROM CAMERA IMAGES

The application described in this section is an example of using our hybrid tracker with a time-invariant system.

There might be instances when a fast reconstruction of the real environment has to be performed because a VR database is not readily available. In such cases, a stereo vision-based technique could be used for building the VR environment from a sequence of 2-D images. In [33], we described a prototype architecture and tool kit, called MIRRORS (Methodology of Inputting Raw Recordings into 3-D Object Renderings for Stereo), designed for performing such tasks. The first version of MIRRORS was based on using a magnetic tracker (Ascension Technologies Flock of Birds) for camera self-calibration. The advantage of employing a tracking system for capturing camera position and orientation is elimination of the necessity of using a calibration pattern, as existing calibration techniques [30], [11] require. Using a magnetic tracker for camera self-calibration limits severely the range of camera motion. On the other hand, a tracking system can calibrate only the extrinsic parameters of the camera while the intrinsic parameters are assumed to remain constant (for a description of camera parameters to be calibrated, the reader is referred to [30], [11], and [33]). As other researchers have also reported [30], [34], assuming some of the intrinsic parameters constant (in particular, the principal point) does not inflict serious errors on depth computation. But the focal length might be a factor on overall errors, especially if the camera undergoes small mechanical or thermal changes that are beyond the control of the user. The contributions of this paper over [33] are usage of the hybrid tracker instead of the magnetic tracker to increase the range of camera motion and the design of a Kalman filter for depth estimation that offers the capability for focal length autocalibration. This technique is, to the best of our knowledge, the only one that enables full-scale reconstruction of 3-D models from stereo images and accurate registration between a virtual and a real environment and is far simpler mathematically and therefore easier to implement than camera self-calibration techniques described in [22] and [34].

For camera motion capture, the hybrid tracker is used in the same way as for human motion capture presented in the previous section. The difference is that the magnetic receiver attached to the PT’s (denoted RT) is used only for the purpose of compensating for the violation of the LOS constraint (provided that the disparity between consecutive images is small—at most two pixels). If this constraint is not violated, the readings collected from PT’s are sufficient for determination of the position and orientation of the camera. The examples given below in this section use only the laser tracker for camera tracking. The camera-tracker unit is depicted in Fig. 7.

In order to retrieve the camera position and orientation w.r.t. WCS, the camera-tracker unit has to be precalibrated in order to determine the invariant transformation between camera and tracker. In this case, as shown in Fig. 2, in addition to calibrating for the transformation TTT (between PT’s and magnetic receiver RT), we need to calibrate for the transformations between CCS (camera coordinate system) and both PT and RT. These transformations have the same meaning as TTT in Fig. 2. For camera-tracker calibration, we still use a regular calibration pattern. Three standard camera calibrations (using the calibration pattern [30]) are required. For simplification, Fig. 2 depicts the case when only two such calibrations (corresponding to positions labeled “a” and “b”) are performed. Concurrently with standard calibration, readings from PT’s and RT are collected. The mathematics of camera-tracker precalibration is similar to the hybrid tracker precalibration, described in Section III, and therefore is omitted here for the interest of brevity. The advantage of using a tracker to calibrate a camera is that it requires a calibration pattern only once, when initializing the system and only for computing the invariant transformation between
camera and tracker. The camera position and orientation within WCS can be found as follows. Consider again Fig. 2. The world-camera matrix at position “i” can be found based on the matrix at position “a” (already known), the relative motion of the PT’s (\(T_{PTab}\)) and the matrix \(T_{TT}\), according to the following equation:

\[
T_{WCb} = T_{TT}T_{PTab}T_{TT}^{-1}T_{WCa}.
\]  

(7)

MIRRORS architecture can be used to register real environments with VE’s created with MIRRORS. This simply implies the computation of the transformation \(T_{WB}\) (as shown in Fig. 2), since the VE is constructed with respect to WCS. Considering the camera position for first image and PT, we can write

\[
T_{WB} = T_{DB}^{-1}T_{TT}^{-1}T_{WCa}
\]  

(8)

thus computing the constant transformation needed to estimate hand position from (1).

A. Depth Estimation

It is well known that in order to compute the WCS coordinates (termed “depth” in computer vision literature) of a point extracted from an object of the visualized scene, at least two images in which the point is visible are needed. In order to improve the accuracy of depth computation, we can use all the images in which the point in question is visible and design a Kalman filter to incorporate this “redundant” information. A Kalman filter has been used before for depth estimation from stereo images [1], [35]. The difference between Kalman filters in these works and our Kalman filter is that we augment the vector of state variables (which normally is the vector of world coordinates of the object points extracted from images \(\mathbf{x} = (x, y, z)^T\) with another variable—focal length of the camera \(f\)). This augmentation enables autocalibration of the focal length without employing intricate algorithms such as the one described in [34]. The world coordinates are computed from the image coordinates of two correspondent points extracted from two images of a stereo pair. We denote the vector of image coordinates as \(\mathbf{z} = (u, v)^T\), the measurement vector for the Kalman filter. The two vectors are related (for a single image) according to the following perspective projection relation:

\[
\tilde{z} = \mathbf{M} \cdot \mathbf{x}
\]  

(9)

where the symbol “tilde” means that the vectors are expressed in projective space and \(\mathbf{M}\) is a \(3 \times 4\) matrix, called perspective projection matrix, whose elements depend on the extrinsic and intrinsic parameters of the camera [10]. By transforming (9) from projective space to Euclidean space, image coordinates \(u\) and \(v\) can be found. The dependence between image coordinates and world coordinates is nonlinear, and therefore the Kalman filter is actually an extended Kalman filter (EKF—see [8] and [24]). The EKF equations for depth estimation and focal length autocalibration are given in Appendix B.

Since the uncertainty in image coordinate measurement is given by the stereo quantization error, whose upper bound is half a pixel [16], the diagonal elements of measurement noise covariance matrix \(\mathbf{R}\) are set to 0.5, and the off-diagonal elements are zero, based on the assumption that the errors in \(u\) and \(v\) are independent. Filtering begins when, for each point of interest extracted from images, the first depth estimate is available through a well-known stereo triangulation method [10], [18]. This estimate serves as the initial value of the state variables representing the world coordinates of the point in question. Also, focal length is initialized with the value obtained when calibrating for the transformation \(T_{TT}\) shown in Fig. 2 (focal length is an output of a standard camera calibration algorithm—[30] and [11]). The error covariance matrix is initialized to some arbitrary value. The variances of the state variables representing world coordinates are computed as follows. The classical stereo triangulation is applied (without Kalman filtering) for a certain number of points extracted from an object. The disparity between consecutive frames is of the same order of magnitude as the one expected when Kalman filtering is used. Ground truth data is also available. The depth computation results are compared with the ground truth, and for each coordinate the mean squared error is considered as the variance in depth computation along that particular coordinate. As for the focal length, multiple camera calibrations using a standard algorithm [30] are performed (without zooming) and the mean squared error of the focal length is considered its variance and supplied to the Kalman filter.

As soon as the point in question becomes visible in another image, the depth is optimized based on standard Kalman filter equations [8], [24]. The Kalman filter with focal length autocalibration is run in parallel for every point extracted from the object. For each point, focal length is initialized with the same value, obtained when precalibrating the camera-tracker unit.

In Fig. 8, an example of applying MIRRORS algorithms to an object is shown. The points of interest that have been extracted from the object are marked, and the correspondent points in the left and right images are numbered accordingly. Also, three epipolar lines are drawn for the purpose of a better illustration.

The best way to test the accuracy in depth computation is to compare the measured dimensions, as resulted from applying MIRRORS, with the physical dimensions of the object. The results are provided in Table III for three cases: without Kalman filtering, Kalman filtering without focal length autocalibration, and Kalman filtering with focal length autocalibration. As can be seen, the Kalman filter improves significantly the accuracy of 3-D measurements, but the focal length autocalibration does not produce a notable improvement. This can be explained by the fact that the internal changes undergone by the camera do not have a big influence on system accuracy.

A more sophisticated example is shown in Fig. 9, which illustrates a manufacturing work-cell reconstructed with MIRRORS and one of the original images. This work-cell is the environment used for human motion tracking.

In Fig. 10, an operator portrayed as an avatar is illustrated operating in a virtual manufacturing environment [the VE is the same as the one shown in Fig. 9(b)]. The human operator (setting up a machine tool) is being tracked with our hybrid tracker, and his motion is being replicated within the VE, \textit{a priori} registered with the real environment, using MIRRORS. The example
shown in Fig. 10 illustrates the potential application of our hybrid tracker to monitoring and training human operators in manufacturing systems.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

An extended-range hybrid tracker for motion capture, based on a combination between a laser tracker and a magnetic tracker, has been described. The use of extended-range tracking in manufacturing environments has been illustrated in human motion tracking and 3-D factory object reconstruction from camera images using a novel camera self-calibration technique.

Possible future improvements can be performed in the following areas.

- A better dynamic model. Even though the dynamic model we have used for human motion capture enables reasonable accuracy, a better dynamic model can still decrease the errors in position estimation. At this time, no mathematical model is available that can precisely describe the nature of human motion. Future research might be able to develop such models.

- Eliminate errors due to different update rates of trackers used as components of the hybrid tracker. An interesting development described in [32] suggests that any incomplete measurement can be incorporated into a Kalman filter as it arrives, without waiting for the complete information required to estimate the state variables. The estimate is recorded only when the complete information is available. We plan to adapt this methodology to our hybrid tracker, by processing separately the information from magnetic and laser trackers and combining them subsequently. This approach can speed up filter prediction, thus reducing the overall latency and improving the accuracy of state estimates.
Fig. 9. Manufacturing work-cell reconstructed with MIRRORS (unwanted details have been removed by the user). (a) One of the original images. (b) Final object.

Fig. 10. A human, operating in a manufacturing system, whose motion is replicated in a VE a priori registered with the real system.

APPENDIX A

KALMAN FILTER FOR HAND AND HEAD TRACKING

As mentioned in Section V, we use constant-acceleration [6], [15] model for tracking the hand position. This model considers slight changes in acceleration as white noise. If we denote the vector of state variables as \( \mathbf{x} \), the discrete-time state equation is

\[
\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{w}(k)
\]

where \( \mathbf{x}(k+1), \mathbf{x}(k) \) is the \( 9 \times 1 \) state vector at time steps \( k+1 \) and \( k \), respectively, \( \mathbf{F} \) is the \( 9 \times 9 \) state transition matrix, and \( \mathbf{w}(k) \) is the \( 9 \times 1 \) process white noise vector, zero-mean and with known covariance matrix \( \mathbf{Q} \). If we consider only the state variables corresponding to one coordinate of motion (say, \( x \)), the state vector looks as follows:

\[
\mathbf{x} = (x, \dot{x}, \ddot{x})^T \tag{A2}
\]

where the velocity and acceleration obey the Newton laws, the state transition matrix (corresponding to \( x \)) looks as follows:

\[
\mathbf{F} = \begin{bmatrix}
1 & \Delta t & \frac{1}{2} \Delta t^2 \\
0 & 1 & \Delta t \\
0 & 0 & 1
\end{bmatrix} \tag{A3}
\]

In (A3), \( \Delta t \) is the time increment between two consecutive observations. In reality, \( \mathbf{F} \) is expanded to accommodate all three coordinates of motion. Process noise covariance matrix \( \mathbf{Q} \) looks as follows (again, only for the coordinate \( x \)—see [6, p. 85]):

\[
\mathbf{Q} = E[\mathbf{w}(k)\mathbf{w}^T(k)] = \begin{bmatrix}
\Delta t^5 \\
\Delta t^3 \\
\Delta t^5 \\
\Delta t^3 \\
\Delta t^5 \\
\Delta t^3 \\
\Delta t^5 \\
\Delta t^3 \\
\Delta t^5
\end{bmatrix} \tag{A4}
\]

In (A4), \( q \) is the constant process noise variance (determined a priori), \( E[\cdot] \) denotes mathematical expectation, and superscript ‘\( ^T \)’ denotes transposition. As with \( \mathbf{F} \), matrix \( \mathbf{Q} \) is expanded to accommodate all three coordinates of motion.

The measurement equation of the Kalman filter is written as follows:

\[
\mathbf{z}(k+1) = \mathbf{H}\mathbf{x}(k+1) + \mathbf{v}(k+1) \tag{A5}
\]

where \( \mathbf{z}(k+1) \) is the \( 3 \times 1 \) measurement vector (hand/head positions along \( x, y, \) and \( z \) axes, as reported by the tracker), \( \mathbf{H} \) is the measurement matrix that relates measurements to the corresponding state variables (in this case, since measurements are also state variables corresponding to position, \( \mathbf{H} = \mathbf{I}_3 \), where \( \mathbf{I}_3 \) is the \( 3 \times 3 \) identity matrix), and \( \mathbf{v}(k+1) \) is the \( 3 \times 3 \) measurement noise vector, zero-mean and with covariance matrix \( \mathbf{R} \) determined a priori.

Based on process and measurement equations, the Kalman filter runs recursively as described in Kalman filtering literature [24], [8], etc.

APPENDIX B

EKF FOR DEPTH COMPUTATION AND FOCAL LENGTH AUTOCALIBRATION

The 4 \( \times \) 1 vector of state variables consists of the world coordinates of points of interest and focal length of the camera \( \mathbf{x} = (x, y, z, f)^T \). Since all four state variables are assumed not to change over time, they are modeled as constants.

The state model is

\[
\mathbf{x}(k+1) = \mathbf{I}_4 \mathbf{x}(k) + \mathbf{w}(k) \tag{B1}
\]

where the indices \( k \) and \( k+1 \) show that the estimate is made at the discrete steps \( k \) and \( k+1 \), respectively, \( \mathbf{I}_4 \) is the \( 4 \times 4 \) identity matrix.
identity matrix, and $\mathbf{w}(k)$ is the process noise vector, zero-mean and with covariance matrix $\mathbf{Q}$. Since the state variables in (B1) are modeled as constants, the matrix $\mathbf{Q}$ is constant throughout the process.

EKF measurement equation is

$$\mathbf{z}(k+1) = h(\mathbf{x}(k+1))$$ (B2)

where $h$ is a function embedding the nonlinear relationship between the measurement and state vector. The function $h$ can be linearized by a first-order approximation of its Taylor series expansion around the current estimate of the state vector. Based on this approximation, the measurement equation is

$$\mathbf{z}(k+1) = \mathbf{H}_{k+1} \mathbf{x}(k+1) + \mathbf{v}(k+1)$$ (B3)

where $\mathbf{H}_{k+1}$ is the Jacobian of $h$ around the current estimate and $\mathbf{v}(k+1)$ is the $2 \times 1$ measurement noise vector, zero-mean and with covariance matrix $\mathbf{R}$. There is one separate Kalman filter for every object point extracted. For each such point, focal length is initialized with the same value, obtained while precalibrating the camera-tracker unit, as described in Section V.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their competent comments. The assistance provided by Industrial Virtual Reality Institute staff and graduate students (in particular, A. Akgunduz and A. Banerjee) is gratefully acknowledged. The authors would also like to thank the Electronic Visualization Laboratory, University of Illinois at Chicago, for assistance with the MotionStar tracking system.

REFERENCES

Pat Banerjee received the M.S. and Ph.D. degrees from Purdue University, West Lafayette, IN, and the B.Tech. degree from the Indian Institute of Technology, Kanpur, India.
He is currently an Associate Professor in the Department of Mechanical Engineering at the University of Illinois at Chicago (UIC), serving as the Director of the Industrial Virtual Reality Institute (IVRI), a joint research and development operation comprising UIC, Northwestern University, and Argonne National Laboratory. His research has been supported by NSF, NIST and ONR, and by companies such as Caterpillar, Searle, and Motorola. He has had a number of papers published in his area of research expertise. His current research interests include virtual reality-based factory design, part design and assembly design models, immersive display interfaces, and linear and nonlinear design optimization problems.
Prof. Banerjee is currently serving as Department Editor of IIE Transactions.

Darren Thompson received the M.S. degree in electrical engineering and computer science in 1998 from the University of Illinois at Chicago and the B.S. degree in computer science from Western Illinois University, Macomb, in 1993.
Currently, he is a Research Engineer for the Industrial Virtual Reality Institute (IVRI). His research interests include virtual humans, tele-collaboration, and the use of real-world devices in virtual reality.