Proceedings of the 5th International Conference on Automation, Robotics and Applications, Dec 6-8, 2011, Wellington, New Zealand

High-Bandwidth Low-Latency Tracking Using Optical and Inertial Sensors

Gontje C. Claasen and Philippe Martin
Centre Automatique et Systèmes, MINES ParisTech
60 boulevard St-Michel, 75006 Paris, France
{caroline.claasen, philippe.martin}@mines-paristech.fr

Frederic Picard
Department of Orthopaedics
Golden Jubilee National Hospital, Glasgow, UK
frederic.picard@gjnh.scot.nhs.uk

***Abstract*—We present an optical-inertial tracking system for tracking a handheld tool in a computer-assisted surgery system. Tracking is fast enough for servo-control of the handheld tool. We show that commercially available optical tracking systems which are suitable for use in computer-assisted surgery systems are not suitable for this application.**

A modified Extended Kalman Filter is used to fuse optical and inertial data. Our approach is a direct one which uses marker image data instead of triangulated position data, thus reducing latencies. Experimental data show that the optical-inertial system can track a moving object at a high bandwidth.

I. INTRODUCTION

Computer-assisted surgery systems have become more and more common in operating rooms. These systems often rely on optical tracking systems tracking both patient's anatomy and surgical tools. An optical tracking system consists of stationary cameras and optical markers attached to the object which is being tracked.

Over the last few years, handheld tools [1], [2] for computer-assisted surgery have been developed. They can potentially save operating time and reduce the amount of surgical material in the operating room compared to mechanical cutting or drilling guides. These tools have to be servo-controlled to keep them in a desired plan. Servo-control demands for a fast tool tracking system which has to be faster than human reaction in order to correct small errors the surgeon might make and to correct deviations caused by a change of bone structure. The tracking system has to have a bandwidth of at least 200Hz and a low latency.

Existing optical tracking systems used in computer-assisted surgery have a low bandwidth of 10-60Hz and an important latency which make them unsuitable for servo-control of a handheld tool.

One possibility to overcome problems of optical tracking is to combine it with another sensor type. An inertial measurement unit (IMU) which consists of accelerometers and gyroscopes is an appropriate choice because it has a high bandwidth, is of small size and low-cost. Inertial sensors cannot be used on their own because of their inherent drifts.

Combining information from both sensor types can be achieved with a so-called data fusion algorithm. It gives an estimation of the object pose (position and orientation) at the higher of the two sensor bandwidths which would be the inertial bandwidth here.

Part of this work was supported by an AXA Research Fund doctoral grant.

For our tool-tracking problem, we are interested in the configuration of a stationary camera and an IMU and optical markers which are rigidly connected and fixed to the object being tracked. Several systems have been presented in the literature which combine optical and inertial data for object tracking. They all use a Kalman Filter to fuse the data from both sensor types.

In [3], the system tracks a pen-like tool to which is attached an IMU and whose tip is detected by four cameras placed on a curved line. The sample rates are 100Hz for the IMU and 20Hz for the cameras. The authors of [3] find that with this setup the position error is reduced compared to optical stereo tracking.

[4] presents a setup consisting of a Vicon optical tracking system with 6 cameras and an optical marker attached to an IMU which is fixed to the object being tracked. With an optical sample rate of 10Hz and the IMU at 100Hz, the system accuracy is found to be as good as the Vicon's at 100Hz.

The authors of [5] combine an optical tracking system with two cameras and three markers with a sample rate of about 55Hz with an IMU being sampled at 500Hz which is attached to the optical markers. They find that pose estimation is possible even during short marker occlusions, as long as at least one marker is visible to the cameras.

In [6], only one camera is used which tracks the position (and the yaw angle when the object is moving) of a fiducial marker pattern which is attached to an IMU being sampled at 400Hz. The camera sample rate is 5 to 10Hz. The authors of [6] report that this system can track a moving object and compensate short-time marker occlusions.

Among the presented systems, [3] and [4] combine optical and inertial data with the goal of reducing the optical sample rate, thus saving processing time compared to optical tracking. They use inertial sensors with a low sample rate of 100Hz and do not use all of their potential.

[5] and [6] use inertial sensors with high bandwidths of 400 and 500Hz, respectively, which would be suitable for servo-control. However, they use marker positions as given by an optical tracking system as measurements (indirect Kalman Filter) which does not solve the problem of their latencies.

Our optical-inertial tracking system has to have a high

978-1-4577-0330-0/11/$26.00 ©2011 IEEE

bandwidth and a low latency to make it suitable for servo-controlling a handheld tool. This defines our goals:

- augment the bandwidth of existing optical tracking systems to 200-300Hz
- make the pose estimation delays as small as possible.

We have developed an optical-inertial tracking system which consists of a stationary stereo camera pair and a sensor unit. The sensor unit is made up of an IMU and optical markers and is attached to a handheld tool. The mathematical model for this system, describing the motion of the sensor unit relative to the cameras and the projection of the markers to the stereo cameras, is presented in Section II.

Section III addresses the data fusion algorithm for the presented system. Estimating the sensor unit pose is achieved by fusing the data from the camera and the IMU with a Kalman Filter. This filter uses a direct approach: The sensor readings are used directly as inputs to the filter. This is opposed to an indirect EKF which uses 3D marker positions (calculated from the marker images by the optical tracking system) as measurements. The direct approach not only reduces latencies by eliminating complex computer vision processing but also prevents the optical tracking errors from being propagated in the filter which should have a positive effect on the accuracy of the pose estimate.

We developed an experimental setup for the optical-inertial system which we describe in Section IV. Using experimental data, we compare our system to purely optical tracking.

II. SYSTEM MODEL

A. Coordinate Systems

The motion of the sensor unit will be expressed in camera coordinates which are denoted by C and are fixed to the right camera center. Their unit vectors are $E_1 = [1,0,0]^T$, $E_2 = [0,1,0]^T$ and $E_3 = [0,0,1]^T$. The camera's optical axis runs along E_1. Image coordinates are expressed in the image sensor coordinate system S which is attached to one of the corners of the camera's image sensor. The left camera coordinate system is denoted by CL and the image sensor coordinate system by SL. The left camera unit vectors are $\tilde{E}_1$, $\tilde{E}_2$ and $\tilde{E}_3$.

The body coordinates, denoted by B, are fixed to the origin of the IMU frame and are moving relative to the camera frames. ${}^B a$ is an acceleration in body coordinates, for example.

Finally, we also use an Earth-fixed world coordinate system, denoted by W.

B. System Dynamics

The following equations describe the sensor unit dynamics:

$$ {}^C\dot{p} = {}^C v \tag{1} $$

$$ {}^C\dot{v} = {}^C G + {}^{BC}q * {}^B a * {}^{BC}q^{-1} \tag{2} $$

$$ {}^{BC}\dot{q} = \frac{1}{2} {}^{BC}q * {}^B\omega \tag{3} $$

where ${}^C G = {}^{WC}q * {}^W G * {}^{WC}q^{-1}$ is the gravity vector expressed in C coordinates. ${}^W G = [0,0,g]^T$ is the gravity vector in the world frame with $g = 9.81 \frac{m}{s^2}$ and ${}^{WC}q$ describes the (constant) rotation from world to camera coordinates. ${}^C p$ and ${}^C v$ are the sensor unit position and velocity, respectively. The orientation of the sensor unit with respect to camera coordinates is represented by the quaternion ${}^{BC}q$. ${}^B a$ and ${}^B\omega$ are the sensor unit accelerations and angular velocities.

C. Output

To project the markers to the camera we use a standard pinhole model [7]:

$$ {}^C y_i = \frac{f}{\langle {}^C m_i, {}^C E_1 \rangle} \begin{bmatrix} \langle {}^C m_i, {}^C E_2 \rangle \\ \langle {}^C m_i, {}^C E_3 \rangle \end{bmatrix} \tag{4} $$

with

$$ {}^C m_i = {}^C p + {}^{BC}q * {}^B m_i * {}^{BC}q^{-1} $$

where y_i is the 2D image of marker i with $i = 1, \ldots, l$ (l is the number of markers). f is the camera's focal distance. ${}^C m_i$ and ${}^B m_i$ are the position of marker i in camera and body coordinates, respectively. $\langle a, b \rangle$ denotes the scalar product of vectors a and b.

The 2D image can be transformed from camera to sensor coordinates by a translation:

$$ {}^S y_i = {}^C y_i + {}^S u $$

where ${}^S u$ is the camera's principal point.

The camera coordinate system in which the sensor unit pose is expressed is attached to the right camera of the stereo camera pair. The transformation between the left and right camera coordinates is expressed by R_{St} and t_{St}:

$$ {}^{CL} p = R_{St} {}^C p + t_{St} \ . \tag{5} $$

D. Noises and Biases

The accelerometer and gyroscope measurements a_m and ω_m of ${}^B a$ and ${}^B\omega$ resp. contain noise and biases:

$$ a_m = {}^B a + \nu_a + {}^B a_b $$
$$ \omega_m = {}^B\omega + \nu_\omega + {}^B\omega_b \ . $$

The accelerometer biases ${}^B a_b$ and gyroscope biases ${}^B\omega_b$ can be represented by a sum of two components: a constant one and a varying one which is modeled as a rate random walk [8]:

$$ {}^B\dot{a}_b = \nu_{ab} \tag{6} $$

$$ {}^B\dot{\omega}_b = \nu_{\omega b} \ . \tag{7} $$

Since the IMU consists of a triad of identical accelerometers and a triad of identical gyroscopes and since we consider the inertial sensor noises to be independent white noises with zero mean, we can write for the auto-covariance

$$ E(\nu_j(t)\nu_j^T(t+\tau)) = \xi_j^2 I_3 \delta(\tau) \tag{8} $$

for $j \in \{a, \omega, ab, \omega b\}$ where δ is the Dirac function.

The marker images are measured in the sensor frame and are corrupted by noise η_y which is also assumed white:

$$ y_m = {}^S y + \eta_y \ . $$

E. Complete Model

The complete model with noises then reads:

$$ {}^{C}\dot{p} = {}^{C}v \tag{9} $$

$$ {}^{C}\dot{v} = {}^{C}G + {}^{BC}q * (a_m - \nu_a - {}^{B}a_b) * {}^{BC}q^{-1} \tag{10} $$

$$ {}^{BC}\dot{q} = \frac{1}{2} {}^{BC}q * (\omega_m - \nu_\omega - {}^{B}\omega_b) \tag{11} $$

$$ {}^{B}\dot{a}_b = \nu_{ab} \tag{12} $$

$$ {}^{B}\dot{\omega}_b = \nu_{\omega b} \ . \tag{13} $$

The outputs for the right and the left camera:

$$ {}^{S}y_i = \frac{f_R}{\langle {}^{C}m_i, {}^{C}E_1 \rangle} \begin{bmatrix} \langle {}^{C}m_i, {}^{C}E_2 \rangle \\ \langle {}^{C}m_i, {}^{C}E_3 \rangle \end{bmatrix} + {}^{S}u_R \tag{14} $$

$$ {}^{SL}y_i = \frac{f_L}{\langle {}^{CL}m_i, {}^{CL}\tilde{E}_1 \rangle} \begin{bmatrix} \langle {}^{CL}m_i, {}^{CL}\tilde{E}_2 \rangle \\ \langle {}^{CL}m_i, {}^{CL}\tilde{E}_3 \rangle \end{bmatrix} + {}^{SL}u_L \tag{15} $$

where $i = 1, \ldots, l$. Indices R and L refer to right and left camera resp. (e.g. f_L is the focal distance of the left camera). The measured outputs are:

$$ y_{im} = {}^{S}y_i + \eta_{yi}, \quad y_{(i+l)m} = {}^{SL}y_i + \eta_{y(i+l)} \ . \tag{16} $$

The six accelerometer and gyroscope measurements $u = [a_m, \omega_m]$ are considered as the inputs of our system and the marker images

$$ y = [\underbrace{{}^{S}y_1, \ldots, {}^{S}y_l}_{{}^{S}y}, \underbrace{{}^{SL}y_1, \ldots, {}^{SL}y_l}_{{}^{SL}y}] \tag{17} $$

as its outputs. y is a vector of length $2l + 2l = 4l$. The state vector which is to be estimated by the data fusion filter in Section III has the form $x = [{}^{C}p, {}^{C}v, {}^{BC}q, {}^{B}a_b, {}^{B}\omega_b]$ and is of dimension 16.

F. Observability

If the system states can be expressed as a function of the inputs, outputs and their derivatives, then the system is observable. The following analysis is done neglecting noise terms.

The ${}^{C}m_i$ can be determined as a function of the output y via triangulation [7]. A solution to the absolute orientation problem, e.g. [9], calculates position ${}^{C}p$ and quaternion ${}^{BC}q$ using the ${}^{C}m_i$. Hence, ${}^{C}p$ and ${}^{BC}q$ can be expressed as a function of the output:

$$ [{}^{C}p^T, {}^{BC}q^T] = \zeta_1(y) \ . \tag{18} $$

According to (9) ${}^{C}v = {}^{C}\dot{p}$ and since ${}^{C}p$ is a function of y,

$$ {}^{C}v = \zeta_2(y, \dot{y}) \ . \tag{19} $$

Equation (10) gives ${}^{B}a_b = a_m - {}^{BC}q^{-1} * ({}^{C}\dot{v} - {}^{C}G) * {}^{BC}q$. Using (18) and (19), this can be written as

$$ {}^{B}a_b = \zeta_3(u, y, \dot{y}, \ddot{y}) \ . $$

Finally, (11) can be transformed to ${}^{B}\omega_b = \omega_m - 2{}^{BC}q^{-1} * {}^{BC}\dot{q}$ which yields, using (18):

$$ {}^{B}\omega_b = \zeta_4(u, y, \dot{y}) \ . $$

All the states can be expressed as a function of the inputs, outputs and their derivatives and hence the system proposed in Section II-E is observable.

G. Quaternions

A quaternion q [10] consists of a scalar $q_0 \in \mathbb{R}$ and a vector $\tilde{q} \in \mathbb{R}^3$: $q = [q_0, \tilde{q}^T]^T$. The quaternion product of two quaternions s and q is defined as

$$ s * q = \begin{bmatrix} s_0 q_0 - \tilde{s}\tilde{q} \\ s_0 \tilde{q} + q_0 \tilde{s} + \tilde{s} \times \tilde{q} \end{bmatrix} . \tag{20} $$

The cross product for quaternions reads $s \times q = \frac{1}{2}(s * q + q * s)$. A unit quaternion can be used to represent a rotation: $R(q)a = q * a * q^{-1}$ where $a \in \mathbb{R}^3$ and $R(q)$ is the rotation matrix associated with quaternion q. If q depends on time, we have $\dot{q}^{-1} = -q^{-1} * \dot{q} * q^{-1}$. If

$$ \dot{q} = q * a + b * q \tag{21} $$

with a,$b \in \mathbb{R}^3$ holds true, then $\|q(t)\| = \|q(0)\|$ for all t.

III. DATA FUSION FILTER

Estimating the pose of the body coordinate system - that is of the sensor unit - relative to camera coordinates is achieved by an extended Kalman filter (EKF). This filter also estimates the sensor unit velocity and the accelerometer and gyroscope biases which are used to correct the IMU data.

A. EKF and Its Variations

1) Extended Kalman Filter: We consider a system

$$ \dot{x} = f(x, u) + M\nu \tag{22} $$

$$ y = h(x, u) + N\eta \tag{23} $$

where ν and η are independent white noises. M is of dimension $n \times r$ and N of dimension $s \times s$ where n is the number of states, r is the number of input noises and s is the number of outputs. The input u and the output y are known. x is the state which is to be estimated. The EKF calculates an estimate $\hat{x}(t)$ of the state $x(t)$ according to

$$ \dot{\hat{x}} = f(\hat{x}, u) - K\big(y - h(\hat{x}, u)\big) \tag{24} $$

$$ \dot{P} = AP + PA^T + Q - PC^T R^{-1} CP \tag{25} $$

where $K = PC^T R^{-1}$, $A = \partial_1 f(\hat{x}, u)$ and $C = \partial_1 h(\hat{x}, u)$. In this notation, ∂_i is the partial derivative with respect to the ith argument. The matrices Q (of dimension $n \times n$) and R (of dimension $s \times s$) depend on $\tilde{Q}$ and $\tilde{R}$ through

$$ Q = M\tilde{Q}M^T \tag{26} $$

$$ R = N\tilde{R}N^T \tag{27} $$

where $\tilde{Q}$ contains the covariance parameters of ν and $\tilde{R}$ those of η (see Section III-C2 for more detail).

The estimation error $\Delta x = \hat{x} - x$ satisfies up to higher-order terms the linear equation

$$ \Delta\dot{\hat{x}} = (A - KC)\Delta x - M\nu + KN\eta \ . \tag{28} $$

Recall that when A and C are constant, e.g. around a steady state point, the EKF converges when (A, M) is stabilisable and (A, C) detectable [11, pp. 704].

2) Multiplicative EKF: When a quaternion has to be estimated, care has to be taken to preserve its unit norm. This is not the case with the standard correction term $K(y-h(\hat{x},u))$ because it does not respect (21). The so-called Multiplicative EKF (MEKF) [12], [13] respects the geometry of the quaternion space and preserves the quaternion unit norm. An error quaternion $e_q = \hat{q}^{-1} * q$ is used instead of q. The correction term for the quaternion differential equation is $\hat{q}K_q(y-h(\hat{x},u))$.

3) Right-Invariant EKF: The idea of the MEKF is developed further in [14] by formulating a new filter called Right-Invariant Extended Kalman Filter (RIEKF). The system and the filter have to be right-invariant w.r.t. a transformation group i.e. applying this transformation to the system equations and the correction terms should leave them invariant. This filter takes into account the system symmetries.

B. Right-Invariant EKF for the Optical-Inertial System

We propose the following right-invariant EKF for the system presented in Section II-E.

1) Filter Equations:

$$
\begin{aligned}
{}^{C}\dot{\hat{p}} &= {}^{C}\hat{v} + K_p E && (29)\\
{}^{C}\dot{\hat{v}} &= {}^{C}G + {}^{BC}\hat{q} * (a_m - {}^{B}\hat{a}_b) * {}^{BC}\hat{q}^{-1} + K_v E && (30)\\
{}^{BC}\dot{\hat{q}} &= \frac{1}{2}{}^{BC}\hat{q} * (\omega_m - {}^{B}\hat{\omega}_b) + K_q E * {}^{BC}\hat{q} && (31)\\
{}^{B}\dot{\hat{a}}_b &= {}^{BC}\hat{q}^{-1} * K_a E * {}^{BC}\hat{q} && (32)\\
{}^{B}\dot{\hat{\omega}}_b &= {}^{BC}\hat{q}^{-1} * K_\omega E * {}^{BC}\hat{q} && (33)
\end{aligned}
$$

with $E = y_m - \hat{y}$. The gain $K = -[K_p, K_v, K_q, K_a, K_\omega]^T$ is calculated according to $K = PC^T R^{-1}$ where P satisfies $\dot{P} = AP + PA^T + Q - PC^T R^{-1} CP$. The calculation of A and C is given in the next section.

The transformation group considered for this filter is

$$
\begin{pmatrix} p^\circ \\ v^\circ \\ q^\circ \\ a_b^\circ \\ \omega_b^\circ \end{pmatrix} := \varphi_{q_0} \begin{pmatrix} p \\ v \\ q \\ a_b \\ \omega_b \end{pmatrix} = \begin{pmatrix} p \\ v \\ q * q_0 \\ q_0^{-1} * a_b * q_0 \\ q_0^{-1} * \omega_b * q_0 \end{pmatrix} .
$$

φ is indeed a transformation group since we have $\varphi_{q_1}\big(\varphi_{q_0}(x)\big) = \varphi_{q_0 * q_1}(x)$. The system (9)–(13) is invariant w.r.t. φ; for example, we have for (11):

$$
\begin{aligned}
{}^{BC}\dot{q}^\circ &= \dot{\overbrace{{}^{BC}q * q_0}} = {}^{BC}\dot{q} * q_0 \\
&= \frac{1}{2}({}^{BC}q * q_0) * q_0^{-1} * (\omega_m - \nu_\omega - {}^{B}\omega_b) * q_0 \\
&= \frac{1}{2}{}^{BC}q^\circ * (\omega_m^\circ - \nu_\omega^\circ - {}^{B}\omega_b^\circ)
\end{aligned}
$$

The filter (29)–(33) is also invariant w.r.t. φ. The transformation group φ represents a quaternion multiplication by the right. This is why the system and filter is called "right-invariant", determining the name of this type of EKF.

2) Error System: For the proposed system we consider the state error $e = [e_p, e_v, e_q, e_a, e_\omega] = [\hat{p} - p, \hat{v} - v, {}^{BC}\hat{q} * {}^{BC}q^{-1}, {}^{BC}q * (\hat{a}_b - a_b) * {}^{BC}q^{-1}, {}^{BC}q * (\hat{\omega}_b - \omega_b) * {}^{BC}q^{-1}]$. These errors are all expressed in camera coordinates. The output error is $E = y_m - \hat{y} = y + \eta_y - \hat{y}$. The error system reads:

$$
\begin{aligned}
\dot{e}_p =& e_v + K_p E \\
\dot{e}_v =& - e_q^{-1} * {}^{BC}\hat{q} * (a_m - {}^{B}\hat{a}_b) * {}^{BC}\hat{q}^{-1} * e_q + \tilde{\nu}_a \\
&+ {}^{BC}\hat{q} * (a_m - {}^{B}\hat{a}_b) * {}^{BC}\hat{q}^{-1} - e_a + K_v E \\
\dot{e}_q =& - \frac{1}{2} e_q * e_\omega + \frac{1}{2}\tilde{\nu}_\omega + K_q E e_q \\
\dot{e}_a =& [e_q^{-1} * {}^{BC}\hat{q} * (\omega_m - {}^{B}\hat{\omega}_b) * {}^{BC}\hat{q}^{-1} * e_q + e_\omega] \times e_a \\
&- \tilde{\nu}_\omega \times e_a - \tilde{\nu}_{ab} + e_q^{-1} * K_a E * e_q \\
\dot{e}_\omega =& [e_q^{-1} * {}^{BC}\hat{q} * (\omega_m - {}^{B}\hat{\omega}_b) * {}^{BC}\hat{q}^{-1} * e_q] \times e_\omega \\
&- \tilde{\nu}_\omega \times e_\omega - \tilde{\nu}_{\omega b} + e_q^{-1} * K_\omega E * e_q
\end{aligned}
$$

where we have set $\tilde{\nu}_j = {}^{BC}q * \nu_j * {}^{BC}q^{-1}$ for $j \in \{a, \omega, ab, \omega b\}$. Notice $\tilde{\nu}_j$ also is a white noise with the same characteristics as ν_j. Indeed, with $R_t = R({}^{BC}q(t))$,

$$
\begin{aligned}
E\big(\tilde{\nu}_j(t)\tilde{\nu}_j^T(t+\tau)\big) &= E\big(\big(R_t \nu_j(t)\big)\big(R_{t+\tau}\nu_j(t+\tau)\big)^T\big) \\
&= R_t E(\nu_j(t)\nu_j^T\big(t+\tau\big)\big)R_{t+\tau}^T \\
&\overset{(8)}{=} \xi_j^2 R_t R_{t+\tau}^T \delta(\tau) = \xi_j^2 I_3 \delta(\tau) \ .
\end{aligned}
$$

We linearize the error system around $\bar{e} = (0,0,1,0,0)$ to bring it into the form (28):

$$
\begin{bmatrix} \delta\dot{e}_p \\ \delta\dot{e}_v \\ \delta\dot{e}_q \\ \delta\dot{e}_a \\ \delta\dot{e}_\omega \end{bmatrix} = (A - KC) \begin{bmatrix} \delta e_p \\ \delta e_v \\ \delta e_q \\ \delta e_a \\ \delta e_\omega \end{bmatrix} - M \begin{bmatrix} \tilde{\nu}_a \\ \tilde{\nu}_\omega \\ \tilde{\nu}_{ab} \\ \tilde{\nu}_{\omega b} \end{bmatrix} + KN\eta_y \ .
$$

We obtain the matrices

$$
A = \begin{bmatrix} 0 & I_3 & 0 & 0 & 0 \\ 0 & 0 & -2[R({}^{BC}\hat{q})(a_m - \hat{a}_b)]_\times & -I_3 & 0 \\ 0 & 0 & 0 & 0 & -0.5 I_3 \\ 0 & 0 & 0 & A_1 & 0 \\ 0 & 0 & 0 & 0 & A_1 \end{bmatrix} \quad (34)
$$

$$
\begin{aligned}
A_1 =& [R({}^{BC}\hat{q})(\omega_m - \hat{\omega}_b)]_\times && (35)\\
C_i =& f_R \frac{1}{\alpha^2} \begin{bmatrix} \beta^2 & -\alpha & 0 \\ \gamma^2 & 0 & -\alpha \end{bmatrix} D && (36)\\
C_{i+l} =& f_L \frac{1}{\chi^2} \begin{bmatrix} \mu^2 & -\chi & 0 \\ \kappa^2 & 0 & -\chi \end{bmatrix} R_{St} D && (37)\\
D =& [I_3 \ \ 0 \ \ -2[R({}^{BC}\hat{q})^B m_i]_\times \ \ 0 \ \ 0] && (38)\\
M =& \begin{bmatrix} 0 \\ diag([-I_3, -0.5 I_3, I_3, I_3]) \end{bmatrix} && (39)\\
N =& I_{4l} && (40)
\end{aligned}
$$

where ${}^{C}\hat{m}_i = [\alpha, \beta, \gamma]^T$ and ${}^{CL}\hat{m}_i = [\chi, \mu, \kappa]^T$. C_i and C_{i+l} each correspond to two lines of the matrix C, associated with marker i. For the cross product, we use the notation $[a]_\times b = a \times b$.

C. Implementation

1) Continuous-Discrete Implementation: We chose a continuous-discrete approach [15] for the implementation since the two sensor types used here have sample rates which differ by a factor of about 5 to 15. The IMU has a high bandwidth and the readings are considered to be continuous whereas the optical readings are regarded discrete.

A prediction step is executed at each new IMU reading. Whenever a camera reading is available, a correction step is done. Since the IMU sample rate is an integer multiple of the optical sample rate, a correction is always preceded by a prediction. Both steps calculate an estimation of the system state: the prediction outputs the so-called a priori estimate $\hat{x}(t)^-$ and the correction the a posteriori estimate $\hat{x}_k$.

2) Filter Tuning: In the continuous-discrete RIEKF, (27) is replaced by

$$R_d = N\tilde{R}_d N^T. \tag{41}$$

where the discrete $\tilde{R}_d$ is related to the continuous $\tilde{R}$ by $\tilde{R}_d = \tilde{R}/\Delta t$ where Δt is the optical sample time.

Matrices Q and R_d are calculated according to (26) and (41) with (39), (40) and with noise covariance parameters for noises ν and η_y resp. which are

$$\tilde{Q} = diag(\xi_a^2 I_3, \xi_\omega^2 I_3, \xi_{ab}^2 I_3, \xi_{\omega b}^2 I_3)$$
$$\tilde{R}_d = \xi_y^2 I_{4l}$$

with

- ξ_a: accelerometer output noise density in $m/s^2/\sqrt{Hz}$
- ξ_ω: gyroscope output noise density in $^\circ/s/\sqrt{Hz}$
- ξ_{ab}: accelerometer rate random walk in $m/s^3/\sqrt{Hz}$
- $\xi_{\omega b}$: gyroscope rate random walk in $^\circ/s^2/\sqrt{Hz}$
- ξ_y: output noise density in pixels.

ξ_a and ξ_ω are given in the IMU datasheet. ξ_{ab} and $\xi_{\omega b}$ are tuning parameters; it is reasonable to increase the datasheet values to take into account the uncompensated bias temperature drift. ξ_y is a tuning parameter as well, including measurement noise and calibration errors.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

Figure 1 shows our sensor unit which is made of an ADIS16355 IMU and three infrared LEDs (four LEDs are visible, but only three are used in the experiments presented here). The IMU is fixed to the back of the sensor unit.

The camera is a Wiimote image sensor which we unsoldered and put on a PCB [16] to use it on its own. The Wiimote is the remote control of Nintendo's game console. This image sensor sees up to four luminous points and outputs the point coordinates via an I2C protocol. We chose this low-cost sensor because it does image processing internally and can easily be interfaced and synchronized with a microcontroller. To build a stereo system, we fixed two image sensors relative to each other in a stereo rig as shown in Fig. 1. Each camera was calibrated using [17] to find its intrinsic parameters using a target with four LEDs in a plane. A stereo calibration was done to find the transformation (5) between the two cameras.

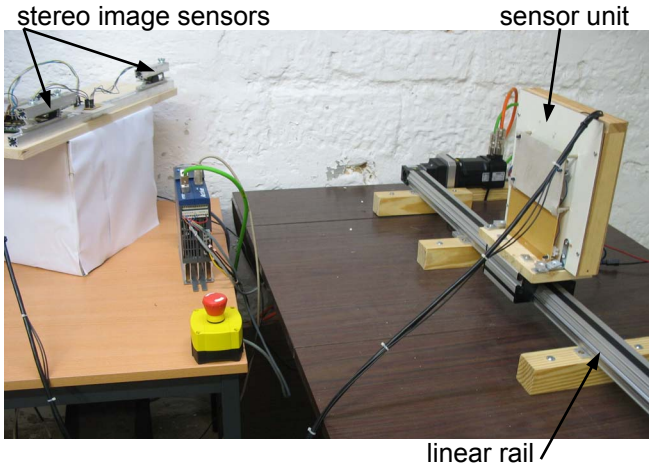

Fig. 1. Experimental setup with stereo rig, sensor unit and linear rail

Note that the size of this experimental sensor unit is much bigger than that of the final sensor unit which would be attached to a handheld tool. This size was chosen to be able to adapt the setup during the development of this experimental setup. Also, the low-cost camera we use demands a distance of several centimeters between the LEDs.

Data from all sensors were acquired by an Atmega2560 microcontroller which communicated with the camera via the Inter-Integrated Circuit (I2C) protocol (we used an I2C switch to change between the two image sensors) and with the IMU via the Serial Peripheral Interface (SPI). The sensor readings were synchronized with a camera sample rate of 25Hz and an IMU sample rate of 250Hz and the data were sent to a PC serial port. These data were processed offline with Matlab/Simulink.

We fed the data to our direct RIEKF which estimated the sensor unit pose. To compare our results to optical tracking, we used the same data to first find 3D marker positions by triangulating [7] the stereo marker image coordinates and then calculated the pose of the IMU center following [9].

We mounted the sensor unit on a linear rail (see Fig. 1). The rail measures the carriage position with a resolver. We used this measurement as a ground truth. In order to compare this 1D measurement to the RIEKF and optical pose estimations, we calculated the position and direction of the rail using RIEKF position estimates and then computed resolver positions relative to this line. In our experiment, the sensor unit executed several back and forth movements on the linear rail.

B. Results

Figure 2 shows results of the experiment we conducted with the experimental setup. The upper subfigure shows the Y coordinate of the position ${}^C p = [X, Y, Z]$, estimated by the RIEKF, by optical tracking and measured by the resolver. The second and third subfigures show the accelerometer and

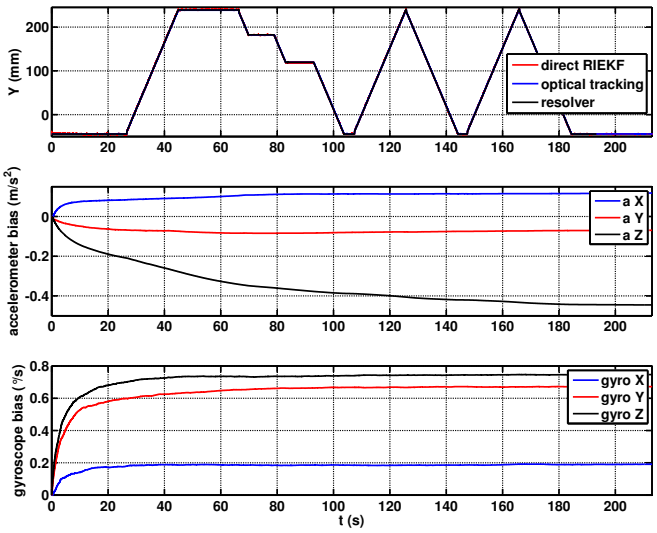

Fig. 2. Position from RIEKF, optical tracking and resolver (top), accelerometer bias (middle) and gyroscope bias (bottom) estimation by RIEKF

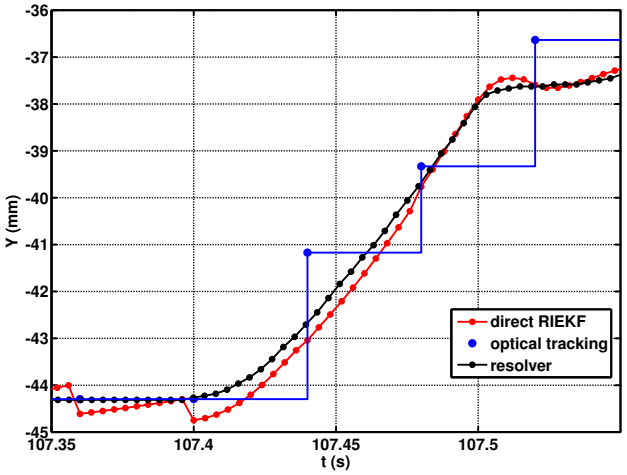

Fig. 3. Detail of Y position coordinate. Note that a constant offset between the RIEKF and optical position exists which is probably due to an imprecise calibration of cameras and sensor unit and which has been compensated here.

gyroscope biases estimated by the RIEKF. They converge which indicates that the RIEKF estimation is correct.

To show the differences between the RIEKF and optical tracking, a detail of the Y position is depicted in Fig. 3. The optical-inertial system with the RIEKF detects the motion at once; the optical tracking detects it with a certain delay.

V. CONCLUSION AND FUTURE WORKS

We presented an optical-inertial tracking system consisting of a stereo camera pair and a sensor unit with an inertial measurement unit and optical markers. This sensor unit can be attached to a handheld tool, i.e. in a computer-assisted surgery system.

We presented the system model and equations for a direct right-invariant Kalman filter estimating position, velocity and orientation of the sensor unit in camera coordinates as well as the accelerometer and gyroscope biases. The motion of the handheld tool is tracked at a high rate of 250Hz and with small latencies which makes servo-control possible.

We developed an experimental setup and conducted an experiment on a linear rail to show that the optical-inertial system with an RIEKF estimates position and biases correctly and to show the benefits of our system over optical tracking using the rail resolver measurement as ground truth. In this experiment, the optical tracking detected the motion later than the optical-inertial tracking. This shows that the optical bandwidth is not high enough to follow fast motion. Note that in a real-time application, optical tracking will have an important latency. This fact was not shown here since processing time was not taken into account in our offline analysis.

Our future work will include a real-time implementation and a more precise calibration of the sensor unit. We also plan to improve the data fusion algorithm to make it even less computationally complex.

REFERENCES

[1] G. Brisson, T. Kanade, A. M. D. Gioia, and B. Jaramaz, "Precision freehand sculpting of bone," in *Medical Image Computing and Computer-Assisted Intervention*, ser. LNCS. Springer, 2004, pp. 105 – 112.

[2] H. Haider, O. A. Barrera, and K. L. Garvin, "Minimally invasive total knee arthroplasty surgery through navigated freehand bone cutting: Winner of the 2005 HAP Paul Award," *J Arthroplasty*, vol. 22, no. 4, pp. 535 – 542, 2007.

[3] N. Parnian, S. P. Won, and F. Golnaraghi, "Position sensing using integration of a vision system and inertial sensors," in *34th Annual Conf. of the IEEE Industrial Electronics Society*, 2008, pp. 3011–3015.

[4] D. Roetenberg, "Inertial and magnetic sensing of human motion," Ph.D. dissertation, Universiteit Twente, 2006.

[5] A. Tobergte, M. Pomarlan, and G. Hirzinger, "Robust multi sensor pose estimation for medical applications," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2009, pp. 492 –497.

[6] B. Hartmann, N. Link, and G. F. Trommer, "Indoor 3D position estimation using low-cost inertial sensors and marker-based video-tracking," in *IEEE/ION Position Location and Navigation Symposium*, 2010, pp. 319 –326.

[7] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2003.

[8] *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros (IEEE Std 95-1997)*, IEEE Aerospace and Electronic Systems Society Std., 1997.

[9] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE T Pattern Anal*, vol. 13, no. 4, pp. 376 –380, 1991.

[10] B. L. Stevens and F. L. Lewis, *Aircraft control and simulation*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2003.

[11] G. C. Goodwin, S. F. Graebe, and M. E. Salgado, *Control System Design*. Prentice Hall, 2001.

[12] E. J. Lefferts, F. L. Markley, and M. D. Shuster, "Kalman filtering for spacecraft attitude estimation," *J Guid Control Dynam*, vol. 5, no. 5, pp. 417–429, 1982.

[13] J. L. Crassidis, F. L. Markley, and Y. Cheng, "Survey of nonlinear attitude estimation methods," *J Guid Control Dynam*, vol. 30, no. 1, pp. 12–28, 2007.

[14] S. Bonnabel, P. Martin, and E. Salan, "Invariant extended Kalman filter: theory and application to a velocity-aided attitude estimation problem," in *48th IEEE Conference on Decision and Control*, 2009.

[15] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[16] U. Jürss and W. Rudolph, "Tracking hot spots," *elektor*, vol. 384, 11 2008.

[17] J.-Y. Bouguet, "Camera calibration toolbox for Matlab." [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc