A Quaternion-based Unscented Kalman Filter for Robust Optical/Inertial Motion Tracking in Computer-assisted Surgery

Nima Enayati, Elena De Momi, and Giancarlo Ferrigno

Abstract— **This paper presents a sensor fusion algorithm based on an Unscented Kalman Filter (UKF) designed for robust estimation of position and orientation of a freely moving target in surgical applications. The UKF is not subject to the nontrivial disadvantages of the more popular Extended Kalman Filter (EKF) that can affect the accuracy or even lead to divergence of the algorithm. Orientation is represented by quaternion which avoids singularities and is computationally more effective. The fusion algorithm has been designed to incorporate an optical tracking system and an inertial sensor unit containing tri-axial angular rate sensors and accelerometers. The proposed tracking system does not suffer from environmental distortion, is robust to brief line-of-sight losses and has a high sampling rate. Experimental results validate the filter design, and show the feasibility of using optical/inertial sensor fusion for robust motion tracking meeting the requirements of surgical computer-assisted procedures.**

Index Terms—**Inertial measurement unit, optical tracking system, pose estimation, quaternion, sensor fusion, surgical navigation, unscented Kalman filter (UKF)**

I. INTRODUCTION

INTRODUCING the idea of frameless and interactive computer-aided surgery provided the surgeons with navigation systems able to show in real time the position of the tip of an instrument in the corresponding pre-operative images without requiring cumbersome implementation of stereotactic frames [1]. This required efficient, fast and accurate motion tracking methods to track the surgical instrument and patient's body parts.

While motion tracking is becoming an inseparable part of modern operating rooms, it has already been a key technology in virtual environments and robotics. A bewildering number of motion-tracking technologies have been designed for different purposes, operating on different physical principles and exhibiting different performance characteristics, including mechanical trackers, acoustic, inertial tracking systems, active magnetic trackers, and optical tracking systems.

Mechanical tracking systems rely on a physical connection between the target and a fixed reference point [2]. The measurement frequency is typically high; however the physical connection constraint imposes a limited range of motion. Acoustic tracking systems emit and sense ultrasonic sound waves to determine the position and orientation of a target, through either time-of-flight or phase-coherence. The system's performance is affected by the changes of environmental factors like temperature, humidity and the distance to the fixed sensors. Magnetic tracking systems can track magnetic field of earth as a widely available DC source to estimate heading. Alternatively, active magnetic tracking systems determine both position and orientation by using orthogonally mounted coils to sense a set of sequentially generated magnetic fields. Electromagnetic tracking systems are responsive and have low levels of latency. However, ferromagnetic and conductive material in the environment can affect a magnetic field's shape and the resulting field distortion substantially affects the accuracy of the tracking. Inertial measurement units (IMU) have become particularly popular due to the availability of low-cost small-size micro-electro-mechanical systems (MEMS). Commonly, these sensors include triads of accelerometers and gyroscopes measuring linear acceleration and angular rate with a relatively high resolution and low latencies. Magnetometers are sometimes included in the unit to add heading estimation capability to the system. IMUs require no external devices or free line of sight and are immune to all forms of interference. However, the drift rates due to integration of the signals are too great for practical long term use [3]. Optical Tracking Systems (OTS) inevitably have two components: light sources and optical sensors. The light sources might be passive objects that reflect or diffuse ambient light or active devices that emit light. Examples of passive light sources include distinguishable fiducials or natural surfaces in the environment and sensors are often multiple CCD cameras. These vision-based tracking systems suffer from a notorious requirement of clear line of sight, end-to-end system delay, and high computational expense. Advantages such as being disposable, having no wiring and wide visibility angle have made infrared retro-reflecting passive markers generally more favorable in surgical applications. However commercial optical tracking systems with passive markers have a lower sampling frequency (20–60 Hz) compared to the ones using active markers (100-800 Hz).

A tracking system employed for surgical applications

All authors are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy (e-mail: nima.enayati@polimi.it).

requires a relatively high degree of robustness, reliability and precision. In general, computer-assisted surgery requires a tracking accuracy in the order of millimeters. 1mm is recommended in computer-assisted neurosurgery [5]. It is not likely to have an emerging sensing technology to solve the problems of current technologies while satisfying the accuracy requirements. As humans and animals have evolved the capability to use multiple senses to improve their ability to survive, a natural solution would be to combine different sensor types to improve the quality, reliability and robustness of the tracking system. Among current solutions, optical tracking systems with passive markers have been widely used in modern operating rooms due to their accuracy and immunity to interference. Combining an OTS with a second sensor may solve or improve the limitations of low frequency of acquisition and clear line-of-sight requirements. Active magnetic tracking systems have been frequently used as a complementary sensor with vision and optical systems since they provide direct position and orientation measurements (compared to inertial measurements that need to be integrated first). However, the sensitivity to surrounding electromagnetic fields and conductive material reduces the reliability of these sensors if used in an ordinary operating room and limits their implementation.

This paper describes the design, implementation, and experimental validation of an Unscented Kalman Filter (UKF) for real-time motion tracking. In order to produce 6D pose estimates relative to an Earth-fixed reference frame, the filter uses input data from an OTS and a sensor module containing a triad of orthogonally mounted linear accelerometers, a triad of orthogonally mounted angular rate sensors, and 4 fixed infrared active markers. Quaternions are used to represent orientation to improve computational efficiency and avoid singularities. The filter increases the low acquisition frequency of the optical tracking system while maintaining the desired accuracy and compensates brief marker occlusion, allowing for an accurate, more robust and faster pose tracking. The UKF uses a set of appropriately chosen weighted points to parameterize the means and covariance of probability distributions. It has been argued [6] that the expected performance of the UKF is superior to that of the EKF and is directly comparable to that of the second order Gauss filter. The differences in performance in some example applications are demonstrated in [7] [8]. A second-order linear system is used for translation and a first order model represents rotation. The filter is experimentally validated using actual sensor measurements in extensive performance tests. The primary contributions of this paper are:

• A UKF designed for robust 6D pose tracking that fuses orientation quaternion input with angular rate measurements and position input with linear acceleration measurements.

• Experimental results validating that filter performance is adequate for surgical applications.

The rest of this paper is organized as follows. An overview of the related work is presented in section II and the approach described in this paper is contrasted with previous approaches. Section III describes the fusion methods including the estimation models. In section IV the experiments are defined, results are presented and discussed. The work is concluded in section V.

II. RELATED WORK

A wealth of research has addressed motion tracking, some of which benefit from multi sensor measurements while others simply implement single sensors to estimate the pose of the tracked object. In an ideal world, a purely inertial tracker would have been an appropriate choice, having advantages of nearly instantaneous measurement, superb resolution and immunity to all forms of interference. However, all the non-idealities such as noise, bias, calibration error and gravity compensation error are accumulated due to the inevitable integration of inertial data, yielding an ever increasing drift. [9] reported a technique to track gravity vector in human motion tracking using merely accelerometers. A set of triaxial accelerometers was used to determine 2D orientation in motions involving small linear acceleration. In case of higher accelerations, two sets of triaxial accelerometers on a single rigid body were used to cancel motion-related linear acceleration. Though the effects of these geometric sensor fusion techniques are depicted, there is no comparison with truth data.

Hybrid systems attempt to compensate for the shortcomings of each technology by using multiple measurements. Numerous approaches have been introduced for motion tracking and registration in augmented reality applications that have tried to compensate the drift by using the so called nine-axis inertial sensor modules, containing three orthogonally mounted triads of angular rate sensors, accelerometers, and magnetometers. Foxlin et al. [10] describe a commercial nine-axis sensing systems designed for head tracking applications. Sensor fusion is performed using a complementary separate-bias Kalman filter (KF). Drift correction is described as only being performed during stationary periods when it is assumed accelerometers are sensing only gravitational acceleration. Thus, the described algorithm requires that the sensors stop moving in order to correct inertial drift errors. More recently, Foxlin et al. [11] developed a 6DOF tracking system with a similar orientation tracking system aided by ultrasonic time-of-flight range measurements of a constellation of wireless transponder beacons. For real-time estimation of rigid body orientation using the Magnetic, Angular Rate, and Gravity (MARG) sensors, Bachmann [12] proposed a non-optimal quaternion-based complementary filter for human body tracking. Extensions to this work and the development of an optimal filter designed for human posture tracking applications are described in [13] and [14]. A Gauss–Newton iteration method is used to preprocess accelerometer and magnetometer data to produce quaternion input to the EKF. As it is mentioned by the authors, the filter is not applicable where accelerations due to forces other than gravity are present for indefinite periods. The comparison of the filter orientation estimates with truth data obtained from a rotary tilt table showed a static accuracy of 2° and a dynamic accuracy of 9°. Kraft [15] describes a quaternion-based UKF for real-time estimation of rigid-body orientation MARG sensors. Simulation results demonstrate the general validity of the described filter.

Performance tests of the filter with real measurements are mentioned, but not shown or quantified. An adaptive filter is proposed in [16] to estimate the external accelerations and to correct them. Ren et al [17] propose a two-stage sensor fusion process for orientation tracking of handheld surgical instruments with MARG sensors. Experiments validate the designed KF for small clinical motion and a small rotational perturbation. Another two-stage Kalman filter was used by [18] in a more application-specific that consists of an orientation estimation step by means of the angular rate measurements and two correction stages to correct the estimated orientation using the acceleration and magnetic field data. While the two-stage MARG sensor approaches satisfy the specified accuracy requirement of 1° and are free of line of sight constraints, the magnetic field distortion can affect the estimation accuracy.

Others have attempted to fuse vision and inertial tracking systems to increase sampling frequency, improve accuracy and add to the robustness. You et al. [19] integrated a natural-feature vision system with three gyro sensors to increase the robustness of vision tracking that provides 3DOF orientation tracking for augmented reality registration. Visual and gyro data were fused using an EKF. This approach was subsequently employed in [20] to perform six-degree-of-freedom object pose tracking. Using only gyro data partially simplified the problem. The experimental results were quite promising in a 2D image plane, but not yet verified in 3D cases. [21] proposed a KF that combines vision tracking and inertial measurements to track a tool. An IMU sampling at 100 Hz is attached to the tool whose tip is detected by four cameras that provide pose estimation at 20 Hz. The authors conclude that using 4 cameras instead of the typical stereo camera vision systems reduces the pose estimation error and integrating vision system with inertial navigation system allows video tracking in lower sampling. Figures show the general validity of the conclusion but the errors are not quantified. [22] used an EKF to fuse marker-based vision tacking and inertial measurement. Unlike the work presented in this paper, the filter is not singularity-free due to the use of Euler angles to represent orientation. The difference between the EKF estimations and vision tracking measurements are reported to be in the order of centimeters in experimental tests while no ground truth comparison is performed.

Active magnetic trackers have been a popular means of tracking in virtual and augmented reality for many years [23]. Even though using magnetic sensors eliminates line-of-sight requirements, they suffer from the susceptibility to distortions of the EM field when placed in close proximity to ferromagnetic objects. [24] proposed a method to decrease the field distortion error by calibrating an electromagnetic tracking system (EMTS) in situ with measurements from an optical tracking system. In this method the EMTS and OTS reference systems are registered once with a calibration procedure, and the pre-computed electromagnetic field distortion correction is then applied to all the subsequent EMTS measurements. This method can compensate for magnetic field distortions, but any change in the arrangement of ferromagnetic objects in tracking system surrounding will cause calibration error unless a new calibration procedure is performed. Vaccarella et al [25] used

an optical/EMTS fusion. Compensation for short marker occlusion was performed in a sampling rate of 10 Hz. Even though the proposed algorithm increases the accuracy of the electromagnetic tracking system in the presence of magnetic field distortion, the susceptibility to distortions of the EM field still remains.

Some work has attempted to lessen the latencies and compensate short marker occlusions of optical tracking systems by implementing inertial sensors. In dynamic registration in augmented reality applications, Azuma and Bishop [4] have used inertial data from accelerometers and angular rate sensors to reduce dynamic errors caused by the end-to-end latency in the position and orientation estimates produced by an outside-in marker based optical tracking system. They have used a predictive pose tracking EKF that guesses where the user's head will be in the future and this has resulted in errors 2-3 times lower than prediction without inertial sensors and 5-10 times lower than using no prediction at all. To compensate for marker occlusion of optical tracking systems, position measurements of individual 3D-markers (tight coupling) at 55Hz were combined by acceleration and angular rate measurements at 500Hz using an EKF by [26]. Errors are sub millimeter as far as one marker is visible. The error reaches 150 mm when all markers are occluded for 3s in experiments. Extension to this work was presented in [27] using an ultra-light coupling algorithm in which measurements from each camera is imported into the filter individually to account for cases where a marker is visible to a camera and not to the others. Another EKF based optical-inertial tracking system was introduced by [28] for a servo-controlled handheld tool in a computer-assisted surgery system. To reduce the latencies marker positions are used directly in the data fusion algorithm (tightly-coupled). To test the performance of the filter the estimation is compared with the measurements of a resolver in linear motions. The errors are not quantified but the results of experiments show the ability of the fusion system to follow the target at an estimation frequency of 250 Hz.

In contrast with the work described above, this paper presents a filter algorithm that is specifically designed for full 6d motion tracking that meets the accuracy requirements of neurosurgery applications. Not being affected by interferences such as magnetic field distortion makes the tracking system a practical choice for operating rooms. The algorithm is singularity-free, computationally efficient for real-time implementations and able to achieve substantially higher sampling frequencies compared to commercial optical tracking systems.

III. SENSOR FUSION METHOD

The Kalman Filter and its extensions have been well investigated in the literature and here only the main concepts are reviewed. In Appendix I a semi-algorithmic description of the implemented UKF is available.

A. The Kalman filter and its extensions

Originally developed for systems assumed to be represented with a linear state-space model, the Kalman algorithm is an optimal state estimator of discrete-time controlled process from

measurements corrupted by normally distributed white noise. The EKF extends the use of the algorithm to systems with non-linear process and measurement models by linearizing the models about the current mean and covariance. A fundamental flaw of the EKF is that the distributions of random variables are no longer normal after undergoing their respective nonlinear transformations and EKF only approximates the optimality of Bayes' rule by linearization. From the intuition that it is easier to approximate a probability distribution than to approximate an arbitrary nonlinear transformation, Julier and Uhlmann [6][7] introduced the unscented transformation (UT) for probabilistic inference. Eliminating the need of derivation and evaluation of Jacobian matrices, the UT-based unscented Kalman filter preserves the normal distributions throughout the non-linear transformations and partially incorporates contributions of higher order information into the estimates.

Among several discrete-time versions of KF the predictor-corrector version seems to be the most commonly used. Given a process that is governed by a non-linear (linear in case of classical KF) stochastic difference equation:

$$x_{k+1} = f(x_k, u_k, w_k) \qquad (1)$$

$$z_k = h(x_k, u_k, v_k) \qquad (2)$$

where the non-linear function f relates the state x and the input u at time step k to the state at the current time step $k+1$ and the non-linear function h relates the state and the input to the measurement z. The random variables w_k and v_k represent the process and measurement noise that will be assumed to be additive here. An a posteriori state estimate is computed as a linear combination of an a priori estimate and a weighted difference between an actual measurement and a measurement prediction. Classical, Extended and Unscented Kalman filter all follow this underlying scheme, yet they do so in a different manner. The UKF addresses the non-linear function approximation issues of the EKF through a fundamentally different approach. It captures the 1st and 2nd order statistics of a random variable that undergoes a non-linear transformation by a minimal set of specifically chosen sample points called sigma points. The sigma points are propagated through the non-linear function (i.e. no linearization is used) to generate a transformed sigma points set. The posterior statistics are found by calculating weighted sample mean and covariance of the transferred sigma points. A summary of the used UKF algorithm is presented at Appendix I. For a more detailed description see [29].

Often in fast pose estimation systems it is desirable to estimate where the object will be in the next step instead of knowing where it is at the moment. This would allow compensating for the possible delays introduced by other systems involved such as robots. In the more common KF the a posteriori state estimate $\hat{x}_k$ is calculated from the measurement z_k. We instead have implemented a predictor-type KF that receives the current measurement z_k and estimates the state for the next step $\hat{x}_{k+1}$. Note that due to presence of a one-step delay between the measurement and the estimation, the performance of the

predictor-type KF is expected to be slightly inferior to that of the normal one. The changes required for this modification in the UKF are explained in Appendix I.
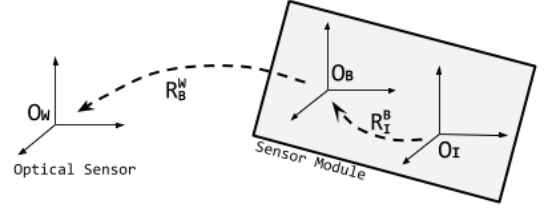

Fig. 1. The reference frames involved in the fusion system

B. Sensor models

The measurements are from multiple sensing modalities that use the following three reference frames:

- World reference frame *W*: The OTS origin coordinate system fixed to the optical sensor and therefore is static and can be assumed as the world reference frame.
- Body reference frame *B*: The OTS dynamic coordinate system fixed to tracked object and defined by the optical markers.
- Inertial reference frame *I*: the coordinate system in which the IMU provides its measurements and is fixed to tracked object.

The OTS reports the pose of the reference frame *B* with respect to its origin coordinate system that is considered as the world reference frame *W* (Fig. 1). Since the measurements of the sensors need to be taken into a single coordinate system for the fusion, the transformation R_I^B between the inertial reference frame *I* and the body reference frame *B* is needed. To find R_I^B a method based on rotation differences explained in [32] was used. The markers and the IMU are rigidly fixed to the sensor module, therefore R_I^B is constant and can be calculated only once. The IMU measurements are taken to the body reference frame before feeding them to the filter. To simplify the notations, here we consider the IMU measurements after the transformation to B reference frame.

In our configuration the optical tracking system is considered as the true signal (ground truth), providing accurate measurements (10 times more accurate than our requirement) of the pose of the rigid body defined by the optical markers. A time delay due to computations is inherent to all loosely coupled tracking solutions. Implementing a tightly-coupled approach that feeds the filter with position measurements of each marker, may reduce the latencies [26]. However in our experimental set the latency of the OTS (approximately 4ms for 4 markers as reported by the manufacturer) was mainly dominated by the communication delays and therefore acquiring marker position would not cause a significant difference. Since the specific latency is relatively small it can be ignored. However, if the latency is significant, care must be taken when fusing these lagged data with the current estimate of the state in the measurement update step of the KF [8]. The IMU as well is assumed to measure without latencies.

The accelerometer and the gyroscope are modeled as:

$$\tilde{a}^B(k) = \bar{a}^B(k) + a_b^B(k) + a_n^B(k) + g^B(k) \qquad (3)$$

$$\tilde{\omega}^B(k) = \bar{\omega}^B(k) + \omega_b^B(k) + \omega_n^B(k) \qquad (4)$$

where $\tilde{a}^B(k)$ and $\tilde{\omega}^B(k)$ are sensor measurements, $\bar{a}^B(k)$ and $\bar{\omega}^B(k)$ are the true values, $a_b^B(k)$ and $\omega_b^B(k)$ are sensor biases, $a_n^B(k)$ and $\omega_n^B(k)$ are measurement noise terms and $g^B(k)$ is the gravitational acceleration, all measured in Body reference frame. The OTS measurements were modeled simply with additive noise terms.

C. Process and measurement models

The output of the measurement system is a set of spatial vectors, measured with respect to sensor reference frame. Given these data, the filter computes an estimate of the system state vector x:

$$x = \begin{bmatrix} p^W & v^W & a_b^B & q & \omega_b^B \end{bmatrix} \quad (5)$$

where $p^W = [x, y, z]$ and $v^W = [v_x, v_y, v_z]$ represent vectors of position and velocity along the world reference frame and $a_b^B = [a_{bx}, a_{by}, a_{bz}]$ and $\omega_b^B = [\omega_{bx}, \omega_{by}, \omega_{bz}]$ are vectors of IMU acceleration and angular rate biases along the body reference frame. $q = [q_0, q_1, q_2, q_3]$ is the orientation quaternion. The system state vector therefore has 16 components. The body's orientation is represented by a unit quaternion. Quaternions offer a singularity-free description of orientation and do not suffer from Gimbal lock as opposed to Euler angles. They are more compact compared to rotation matrices and thereby are more computationally efficient. The set of unit quaternions i.e. a quaternions of norm one, constitutes a unit sphere in four-dimensional space. The three remaining degrees of freedom after applying unity constraint are enough to represent any rotation in 3D space.

Knowledge of the state allows theoretical prediction of the future (and prior) dynamics and outputs of the deterministic system in the absence of noise. The continuous-time kinematic equations that govern the state vector are:

$$\dot{p}^W = v^W \quad (6)$$

$$\dot{v}^W = R_B^W(\tilde{a}^B - a_b^B) - g^W \quad (7)$$

$$\dot{q}^W = \frac{1}{2} q^W \otimes (\tilde{\omega}^B - \omega_b^B) \quad (8)$$

$$\dot{a}_b^B = w_{a_b} \quad (9)$$

$$\dot{\omega}_b^B = w_{\omega_b} \quad (10)$$

where vectors w_{a_b} and w_{ω_b} are zero mean Gaussian variables and $\otimes$ represents quaternion multiplication. For a thorough description of quaternion algebra refer to [31]. Since the vector of the gravitational field g is constant in the world reference frame, by converting the acceleration measurements from body to world reference frame, gravity can be removed from the measured accelerations. This is performed in the process model (7) where g^W is the gravitational acceleration vector in the world reference frame and R_B^W is the direction cosine matrix (DCM), a nonlinear function of the current orientation unit quaternion that transforms vectors from body reference frame to the world reference frame given by:

$$R_B^W = 2 \begin{bmatrix} 0.5 - q_2^2 - q_3^2 & q_1 q_2 - q_0 q_3 & q_1 q_3 + \\ q_1 q_2 + q_0 q_3 & 0.5 - q_1^2 - q_3^2 & q_2 q_3 - \\ q_1 q_3 - q_0 q_2 & q_2 q_3 + q_0 q_1 & 0.5 - q_1^2 \end{bmatrix} \quad (11)$$

Equation (7) is a nonlinear function of the state variables q and a_b^B. The translation discrete time-updates are calculated by the following simple second order Euler update:

$$p_{k+1}^W = p_k^W + \dot{p}_k^W.dt + \frac{1}{2}\dot{v}_k^W.dt^2 \quad (12)$$

$$v_{k+1}^W = v_k^W + \dot{v}_k^W.dt \quad (13)$$

$$a_{b_{k+1}}^B = a_{b_k}^B \quad (14)$$

where dt is the length of the time interval. $\dot{p}_k^W$ and $\dot{v}_k^W$ are calculated using (6) and (7). The accelerometer and gyroscope measurements are used as the inputs of the filter.

The discrete quaternion propagation is described by a first order update equation i.e. constant angular velocity considering that a rotation by q_1 followed by rotation by q_2 is equivalent to a rotation by $q_2 \otimes q_1$. The update equations therefore are:

$$q_{k+1} = q_k \otimes q_d \quad (15)$$

Having the input ω_k^W and the length dt of the time interval, the differential rotation during this interval q_d is defined as:

$$q_d = \begin{bmatrix} \cos\frac{\alpha_d}{2} & , e_d \sin\frac{\alpha_d}{2} \end{bmatrix} \quad (16)$$

where α_d and an axis e_d are angle and axis of rotation:

$$\alpha_d = |\omega_k^W|.dt \quad (17)$$

$$e_d = \frac{\omega_k^W}{|\omega_k^W|} \quad (18)$$

$|\omega_k^W|$ is the total angular velocity in rad/s. The measurement model projects the measurement data onto the state estimate and describes the influence of a random measurement noise on the measured value. With the specific defined state vector, a simple linear model can be used as the measurement model for the implemented sensors:

$$z_{p_k} = p_k + v_{p_k} \quad (19)$$

$$z_{q_k} = q_k + v_{q_k} \quad (20)$$

where v_{p_k} and v_{q_k} are the stochastic noise terms for the OTS measurements of position and orientation.

D. Noise covariances

The process noise covariance matrix Q_k that represents the uncertainty of the system state estimate is defined by:

$$Q_k = E[w_k w_k^T]$$

$$w_k = \begin{bmatrix} w_{p_k}, w_{v_k}, w_{a_{b_k}}, w_{q_k}, w_{\omega_{b_k}} \end{bmatrix}$$

where E is the expectation operator and $\boldsymbol{w}_k$ is the discrete process noise vector of (22). Even though there is no necessary condition on diagonality of $\boldsymbol{Q}_k$, it is a standard assumption for practical purposes to set it as a diagonal matrix. However, the unity constraint on quaternions makes the quaternion components of the state vector dependent and therefore the process noise covariance matrix will not be accurately defined by a diagonal 4 by 4 matrix. Since the orientation component of the state vector has in fact three degrees of freedom the corresponding process noise can be described with a three-dimensional noise vector $\boldsymbol{w}_{q_k}$. The resulting $\boldsymbol{Q}_k$ is a 15-dimensional diagonal matrix. The level of "trust" on the accuracy of the measurements is represented by measurement noise covariance matrix $\boldsymbol{R}_k$ defined as:

$$\boldsymbol{R}_k = E\left[\boldsymbol{v}_k \boldsymbol{v}_k^T\right] \tag{23}$$
$$\boldsymbol{v}_k = \left[\boldsymbol{v}_{p_k}, \boldsymbol{v}_{q_k}\right] \tag{24}$$

The component that represents the measurement noise on orientation $\boldsymbol{v}_{q_k}$ is defined as a 3-dimensional vector for the same reason that was described for process noise variable. Noise characteristics of the position measurements were directly determined from various static acquisitions. For the orientation component the quaternion measurements were first converted to rotation vector and the 3-dimensional noise variance was then calculated.

A subtlety of the filter is the mismatch of the dimension of noises matrices and state vector. In all algebraic operations the quaternion part must be considered separately. As an example in the calculation of sigma points where a set of disturbed state vectors are generated by adding noise to the state, the noise corresponding to the orientation quaternion is first converted to a quaternion and then the state vector is quaternion-multiplied by the generated quaternion. Another subtlety is that when the mean of the transformed sigma point is calculated it must be taken into account that orientations are periodic and an arithmetic mean may achieve wrong results. We have implemented the intrinsic gradient descent algorithm that is described in [32] to find the mean value of the quaternion sets.

IV. EXPERIMENTAL RESULTS

The performance of the developed fusion algorithm was put to test by acquiring real data from an experimental setup consisting of:
- An optical tracking system
- A sensor module: the board to which the inertial measurement unit and 4 active optical markers were fixed
- A computer running Linux Ubuntu operating system, in charge of data storage, implementing the packages that we developed in C++ for Robotic Operating System (ROS).

The OTS is Optotrak Certus (Northern Digital - Canada) that uses 3 charge-coupled devices (CCD) to obtain 3D position of infrared light emitting diodes within a working volume. The device provides 6DOF pose measurement in case of having a rigid body defined by at least 3 markers. The tracking frame rate depends on the number of markers. For instance in case of having 4 markers the maximum reachable frequency of tracking

is claimed to be about 850 Hz. However in practice the maximum achieved real-time frequency for 4 markers was 300Hz. The IMU is Inertial Two sensor module from ATMEL (Atmel, San Jose, USA). It delivers a nine degree-of-freedom sensor platform including a 3-axis accelerometer (KXTF9-1026 from Kionix), a 3-axis compass (HMC5883L from Honeywell) and a 3-axis gyroscope (IMU-3000 from InvenSense). The sensors provide various sampling ranges (max: 2000 °/s for Gyro and 8 g for accelerometer) and frequencies (max: 2100 Hz for Gyro and 800 Hz for accelerometer). However, the final sampling frequency depends on the performance of the microcontroller board and the quality of communications. The IMU is mounted on the evaluation board (UC3-A3 Xplained) with Atmel AT32UC3A3256 32-bit AVR microcontroller. The IMU and 4 optical markers are rigidly connected and fixed to the sensor module as it can be seen in Fig. 2. Software packages for data acquisition from the sensors were developed in C++ implementing the modular architecture of ROS framework. A firmware was written for the IMU's microcontroller for serial communication. The fusion algorithm was developed in MATLAB R2010b, and it was run offline by having the recorded data acquisitions fed to it.

The performed experiments fall into two categories: frequency augmentation and marker occlusion. In both tests the sensor fusion algorithm estimates the sensor module's pose while the OTS measurements are not available due to either a lower sampling frequency (frequency augmentation test) or temporary loss of marker sights (marker occlusion test). In each category, the experiments were performed separately for orientation and position estimation. The sensor module was moved by hand in space. The details of each experiment category are described next.

Fig. 2. The optical tracking system and the sensor module used for the experiments

A. Frequency Augmentation Experiments

Existing commercial optical tracking systems used in computer-assisted surgery have a bandwidth of 20–60 Hz. In order to evaluate the performance of the fusion algorithm for frequency augmentation purposes, we fed the algorithm with a down-sampled 20Hz OTS measurement and a 200Hz IMU measurement. The frequency of the estimated pose by the fusion algorithm is equal to that of the faster input i.e. the 200Hz IMU signal. Since between the updates, the OTS measurement is not used by the fusion algorithm, the accuracy of the pose estimation can be found by comparing the estimated pose with the ground truth signal that is the OTS measurement at 200Hz. We define the frequency augmentation (FA) ratio as the ratio of the fusion algorithm's output frequency to that of the OTS. For example having an FA ratio of 10 means that for each 10 estimated samples from the IMU, 1 sample is acquired from the OTS that "corrects" the drift caused by estimation from the

IMU. The lower end of the frequency range of OTS devices (20 Hz) was chosen since it represents the worst case. The higher the FA ratio is the longer the IMU estimation interval will be, which results in higher drifts. The architecture of the frequency augmentation test is shown in Fig. 3.
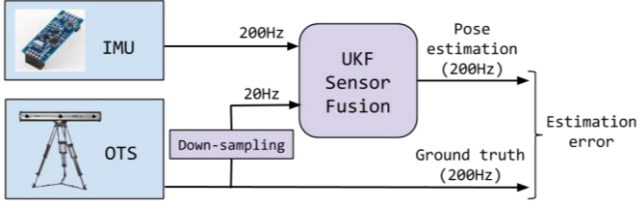

Fig. 3. Frequency augmentation experiment scheme

In position estimation experiments the sensor module was moved for 1 minute in a circular trajectory (xy plane in world reference frame) with an approximate diameter of 30cm. The orientation experiments involved a periodic rotation of about 60 degrees around *y* axis in body reference frame. In motion tracking systems the estimation error depends on dynamics of the motion. The higher the acceleration and velocity of the object the greater will be the estimation error. [33] reports the maximum velocity of patient's head movement during awake brain surgery to be 60 mm/s. To account for cases involving higher velocities, the position experiments were divided in two velocity ranges with RMS linear velocity ranges of 35-75mm/s and 100-150mm/s. For orientation experiments a single angular rate set ranging 40°/s to 60°/s RMS was performed. For each set of experiments 12 repetitions were recorded. The estimation error was measured at each step by comparing the estimated pose with the real pose (OTS measurement). The average of the 12 root mean square errors (RMSE) of the 12 repetitions for each experiment set was computed as the performance of the fusion algorithm.

TABLE I SUMMARY OF THE PERFORMED EXPERIMENTS. P AND O STAND FOR POSITION AND ORIENTATION.

		Sensor Frequency (Hz)		Velocity Range (RMS)	Repetitions
		OTS	IMU		
Frequency Augmentation	P	20	200	35-75 mm/s	12
		20	200	100-150 mm/s	12
	O	20	200	40-60 °/s	12
Marker Occlusion	P	100	200	35-70 mm/s	12
	O	100	200	8-12 °/s	12

B. Marker occlusion Experiments

During marker occlusion, the filter estimates the pose from IMU measurements and the error due to noises and discrete integration is accumulated. The longer the occlusion time the greater will be the error. Marker occlusion experiments evaluate the performance of the algorithm when it is being used to compensate for temporary loss of marker visibility. In these experiments the sensor module was moved/rotated nonstop in random trajectories with various velocities for 15s and data were acquired from both sensors at 200Hz. To be able to measure the estimation error during the occlusion, the markers were not covered during the acquisitions. Instead, the occlusions were simulated by not feeding the OTS measurements to the fusion algorithm during the occlusion periods. As shown in Fig. 4, an occlusion window lasting 200

samples was applied N times (N = 14), shifting it over the entire acquisition. 12 acquisitions were performed and 14 occlusions were simulated per acquisition resulting in a total of 168 occlusions. To evaluate the performance of the algorithm as function of the occlusion duration, the errors during each occlusion is recorded and the root mean square error (RMSE) of all the 168 occlusions is calculated as a function of the sample progression from the beginning of the occlusion window as:

$$RMSE_{MO}(k) = \sqrt{\frac{1}{12N}\sum_{i=1}^{168}\left(\hat{x}_k^i - x_k^i\right)^2} \qquad (25)$$

where k is the sample number (0-200). $\hat{x}_k^i$ and x_k^i are the estimated and real pose of occlusion I at sample k. Details of the experiments are summarized in Table I.
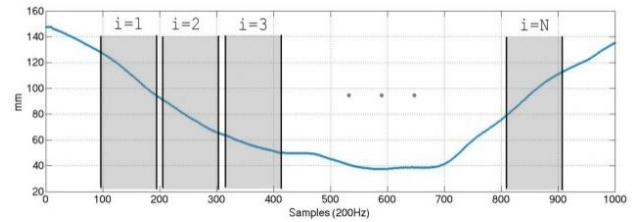

Fig. 4. In marker occlusion experiments a fixed occlusion window is shifted over the signal for N times and the estimation error during each occlusion is recorded.

C. Results of the Frequency Augmentation Experiments

Fig. 5 shows the estimated and real (reference) position trajectories of the sensor module in a test acquisition. It can be seen from the enlarged part of the drawing that the estimation has a saw-tooth shape since the estimated trajectory drifts away from the reference trajectory until the OTS sample corrects the drift. Clearly the sensor fusion system is capable of tracking the sensor module. 5.5s of estimated and reference orientations around axis y along with the estimation error in an FA orientation estimation test is shown in Fig. 6. The error signal shows the saw-tooth drift correction phenomenon in a clearer manner. Furthermore, it can be observed that the maximum errors (tips of the saw-tooth) have a sine-shaped trend that follows the trend of the derivative of the estimated orientation. In fact, the Kruskal-Wallis test on the population of the RMSE values of the lower velocity set and the population of the higher velocity set yielded a p value of 0.003 confirming the increase of the RMS errors with the increase of velocity. The results are summarized in Table II. The reported velocities are the ranges of the 12 acquisitions' RMS velocity. From the experiments we observe that the proposed system is able to yield pose estimation at 200Hz with a maximum RMS errors of 0.57mm and 0.43° using 20Hz optical measurement and 200Hz Inertial measurement. The corresponding 3D position tracking accuracy is 0.37mm for the lower velocity range and 0.75mm for the higher velocity range.

TABLE II FREQUENCY AUGMENTATION RESULTS SUMMARY. RMSE OF ESTIMATION ERROR IN FA EXPERIMENTS

		Velocity range (RMS)	RMSE
Position Velocity Range 1	x	35-50 mm/s	0.17mm
	y	55-75 mm/s	0.29mm

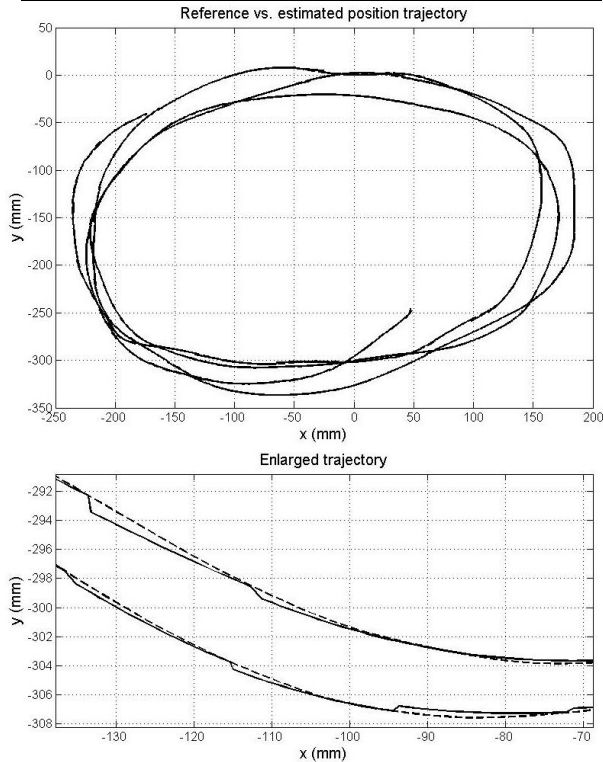|			
	z	15-20 mm/s	0.16 mm
Position Velocity Range 2	x	100-130 mm/s	0.42mm
	y	100-150 mm/s	0.57mm
	z	20-35 mm/s	0.22mm
Orientation	x	2-8 °/s	0.29 °
	y	10-20 °/s	0.32 °
	z	50-60 °/s	0.43 °

Fig. 5. Real position (dashed line) and the estimated position (solid line) for an acquisition with FA ratio of 10. Lower figure shows an enlarged part of the upper one.
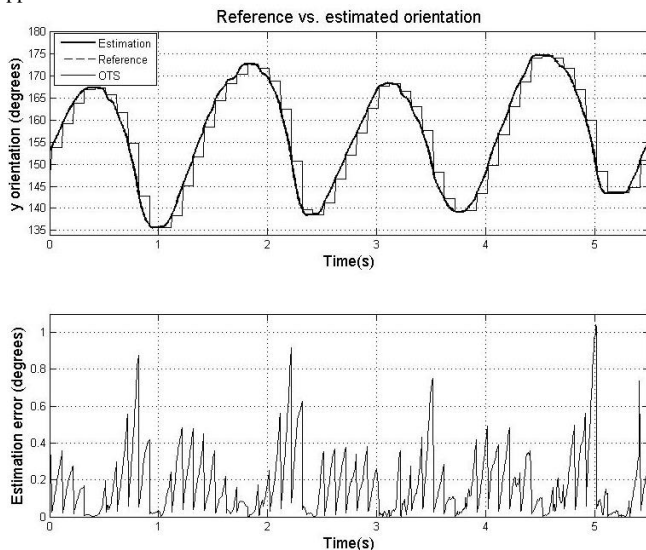

Fig. 6. Up: Orientation estimation using 20Hz OTS and 200Hz Gyroscope measurements. The absolute error between the estimated orientation and reference (OTS at 200Hz) is shown in the lower plot.
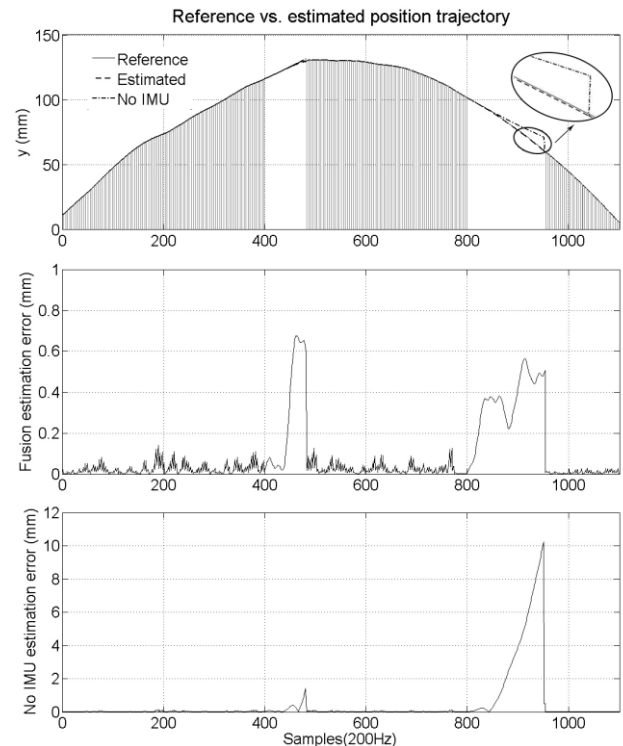

Fig. 7. Position estimation using inertial measurements during 2 simulated marker occlusions. Vertical bars represent the availability of OTS measurement. Lower plots shows the estimation error at each step for Fusion and no IMU estimations.

D. Marker Occlusion Results

Fig. 7 depicts how the fusion system responds to marker occlusion. In the example acquisition shown, two occlusions with durations of 80 and 150 samples happen. As it can be seen the object has been moved about 130mm and returned to the first position with an average velocity of approximately 45mm/s. The vertical lines represent the availability of OTS measurements. The middle plot in Fig. 7 shows that during the occlusions the estimation error increases and gets corrected as soon as OTS measurement revives. To show the improvement achieved by fusing the inertial measurement a third signal has been added to the plot that is position estimated by the UKF with a second-order model using only the last available OTS measurement and no IMU input. As it can be seen the error reaches 10mm in the second occlusion while for fusion the error stays below 1mm. This example acquisition however, is from a smooth motion. In case of highly dynamic and random motions the errors can be higher. To evaluate the performance of the system the RMS error of 168 marker occlusions during random motions are plotted as a function of duration of occlusion in Fig. 8. It can be observed that the error increases by extending the duration of markers occlusion reaching a maximum of 0.88° and 2.78mm at the end of 200 samples of estimation. The position estimation error stays under 1mm for a maximum of 60 samples while the error of the orientation does not exceed the 1° limit during the occlusion period. The user must be informed if the duration of the occlusion passes the threshold for the required accuracy.
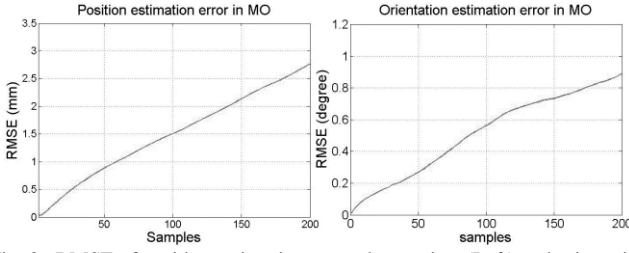
Fig. 8. RMSE of position estimation error along axis y (Left) and orientation estimation error around axis y for 54 simulated occlusions

V. CONCLUSION

In this work a quaternion based UKF was developed to be used as a fusion algorithm for Optical and Inertial motion tracking in medical surgery applications. Despite that the problems of marker occlusion and frequency augmentation for optical tracking systems have been previously addressed by researches, the accuracy of those efforts were either insufficient or not evaluated. Commercial optical tracking systems used in OR are almost all based on passive markers and have a relatively low sampling frequency. The presented fusion algorithm employs an inertial measurement unit to augment the measurement frequency of an optical tracking system, yielding a tracking system 10 times faster than the OTS. The accuracy of the system was evaluated under extensive experimental tests for different ranges of velocities and proved to be below 1mm and 1° that satisfies requirements of tracking in computer-assisted neurosurgery. In addition, the tracking system is robust to brief loss of sight of the optical markers, a case in which a purely optical system would fail to provide accurate measurements. The developed sensor fusion system is based on quaternion representation for orientation; an advantage compared the common Euler angles that suffer from kinematic singularities. In this work the fusion algorithm was used offline, a future work will be taking into account the latencies in the fusion algorithm and developing a real-time tracking platform. Another future development can be studying the frequency of occurring partial marker occlusion compared to that of total occlusion in an operation. In case of significant difference a tightly coupled algorithm can be implemented that can possibly improve the estimation accuracy during partial marker occlusion.

VI. APPENDIX I: UKF ALGORITHM

This section provides a brief description of the UKF algorithm used in the fusion system. The algorithm is based on the scaled UKF [29] with a few minor modifications.

A. Initialization

In the beginning of the recursive algorithm initial values for the state vector x_0 and state covariance matrix P_0 is assumed. A typical choice is the first sensor measurements for the measured states and zero for the unmeasured ones. The initial covariance matrix can be set as an identity matrix.

B. Sigma point generation

For a random variable x of dimension n with mean $\bar{x}$ and covariance P_{xx}, $2n + 1$ "scaled sigma points" and their corresponding weights are given by:

$$\mathcal{X}_0 = \bar{x} \qquad W_0^m = \frac{\lambda}{(n + \lambda)} \qquad (26)$$

$$\mathcal{X}_i = \bar{x} + (\sqrt{(n + \lambda)P_{xx}})_i \qquad W_0^c = \frac{\lambda}{(n+\lambda)} + (1 - \alpha^2 + \beta)$$
$$i = 1, \dots n$$

$$\mathcal{X}_i = \bar{x} - (\sqrt{(n + \lambda)P_{xx}})_i \qquad W_i^m = W_i^c = \frac{1}{2(n+\lambda)}$$
$$i = n, \dots 2n \qquad i = 1, \dots 2n$$

where $(\sqrt{(n + \lambda)P_{xx}})_i$ is the ith row or column of the matrix $\sqrt{(n + \lambda)P_{xx}}$. A Cholesky decomposition can be used to compute $\sqrt{(n + \lambda)P_{xx}}$ since P_{xx} is a symmetric and positive definite. λ is equal to:

$$\lambda = \alpha^2(n + \kappa) - n \qquad (27)$$

Parameters κ, α and β determine the spread of the sigma points around the prior mean [29]. For readability subscript k was omitted from the sigma points $\mathcal{X}_k$.

C. Time-Update equation

The sigma points are propagated through the process model to get the transformed sigma points $\mathcal{Y}_k$:

$$\mathcal{Y}_{k+1|k} = A(\mathcal{X}_k, u_k) \qquad (28)$$

The predicted mean $\hat{x}_k^-$ (*a priori* estimate) can be computed as:

$$\hat{x}_{k+1|k}^- = \sum_{i=0}^{2n} W_i^m \mathcal{Y}_i \qquad (29)$$

The predicted covariance $P_{y_k}^-$ is calculated as:

$$P_{y_{k+1}}^- = \sum_{i=0}^{2n} W_i^c (\mathcal{Y}_{i,k+1|k} - \hat{x}_{k+1}^-)(\mathcal{Y}_{i,k+1|k} - \hat{x}_{k+1}^-)^T \qquad (30)$$

D. Measurement-update equation

In normal UKF the transformed sigma points are propagated through the measurement model as:

$$\mathcal{Z}_{k+1|k} = H(\mathcal{Y}_{k+1}) \qquad (31)$$

Since we aim to implement a predictor type KF, we propagate the sigma points instead of the transformed set:

$$\mathcal{Z}_{k|k} = H(\mathcal{X}_k) \qquad (32)$$

The predicted observation is calculated as:

$$\hat{z}_{k|k}^- = \sum_{i=0}^{2n} W_i^m \mathcal{Z}_{k|k} \qquad (33)$$

The innovation covariance matrix can be calculated next:

$$P_{z_k} = \sum_{i=0}^{2n} W_i^c (\mathcal{Z}_{k|k} - \hat{z}_k^-)(\mathcal{Z}_{k|k} - \hat{z}_k^-)^T \qquad (34)$$

The cross-correlation covariance matrix is:

$$P_{x_{k+1}z_k} = \sum_{i=0}^{2n} W_i^c (\mathcal{Y}_{i,k+1|k} - \hat{x}_{k+1}^-)(\mathcal{Z}_{k|k} - \hat{z}_k^-)^T \quad (35)$$

And finally the Kalman gain is calculated as:

$$K_{k+1} = P_{x_{k+1}z_k} P_{z_k}^{-1} \quad (36)$$

E. The a posteriori estimates

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(z_k - \hat{z}_k^-) \quad (37)$$

$$P_{y_{k+1}} = P_{y_{k+1}}^- - K_{k+1} P_{z_k} K_{k+1}^T \quad (38)$$

REFERENCES

[1] P. B. McBeth, D. F. Louw, P. R. Rizun, G. R. Sutherland. Robotics in neurosurgery, *American J. Surgery* 188 pp. 68S-75S. Oct. 2004.

[2] Products: Force Dimention, Sigma.7DOF2013 [Online]. Available: http://www.forcedimension.com/sigma7-overview

[3] G. Welch and E. Foxlin, Motion tracking: no silver bullet, but a respectable arsenal, *IEEE Comput. Graphic. Applicat.*, vol. 22, no. 6, pp. 24–38, Nov.–Dec. 2002.

[4] R. Azuma and G. Bishop, Improving Static and Dynamic Registration in an Optical See-through HMD, *Proc. of SIGGRAPH* 95, 1995.

[5] M. D. Comparetti, E. De Momi, T. Beyl, M.Kunze, J. Raczkowsky, and G. Ferrigno, Convergence analysis of an iterative Targeting Method for Keyhole Robotic Surgery, *Intern. Jour. Advanc. Robot. Sys.*2014.

[6] S. J. Julier and J. K. Uhlmann, A New Extension of the Kalman Filter to Nonlinear Systems, Proc. SPIE - Int. *Soc. Opt. Eng. (USA)*, vol. 3068, pp. 182-193, April 1997.

[7] S. J. Julier and J. K. Uhlmann, Unscented filtering and nonlinear estimation, *Proc. of IEEE*, vol. 92, pp. 401-422, Mar 2004.

[8] R. van der Merwe and E. Wan. Sigma-point Kalman filters for integrated navigation. *Proc. Annu Meet. of the Inst. of Navig, Dayton, OH*, Jun. 2004.

[9] J. Lee and I. Ha, Sensor fusion and calibration for motion captures using accelerometers, *Proc. IEEE Int. Conf. Robot. Autom.*, *Detroit, MI*, pp. 1954–1959, May 1999.

[10] E. Foxlin, Inertial head-tracker fusion by a complementary separate-bias Kalman filter, *Proc. Virt. Real. Annu. Int. Symp., Santa Clara, CA,* pp. 185–194. Mar. 1996.

[11] E. Foxlin, M. Harrington, and G. Pfeifer. Constellation: A Wide-Range Wireless Motion-Tracking System for Augmented Reality and Virtual Set Applications, *Proc. Annu. Conf. Comput. Graph. and int. techn,* pp 371-378 98, 1998.

[12] E. R. Bachmann, Inertial and magnetic tracking of limb segment orientation for inserting humans into synthetic environments, *Ph.D. dissertation, Naval Postgraduate School, Monterey, CA*, 2000.

[13] X. Yun, E. Bachmann, Design, implementation, and experimental results of a quaternion-based kalman filter for human body motion tracking. *IEEE Trans Robot* 22(6):1216–1227, 2006.

[14] X. Yun, E. Bachmann and R. B. McGhee, "A Simplified Quaternion-Based Algorithm for Orientation Estimation From Earth Gravity and Magnetic Field Measurements," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 3, pp. 638–650, Mar. 2008.

[15] E. Kraft, A quaternion-based unscented Kalman filter for orientation tracking, *Proc. IEEE 6th Int. Conf. Inf. Fusion,* Cairns, Queensland, Australia, pp. 47–54, 2003.

[16] S. Suh, "Orientation estimation using a quaternion-based indirect Kalman filter with adaptive estimation of external acceleration*," IEEE Trans. Instrum. Meas.*, vol. 59, no. 12, pp. 3296–3305, Dec. 2010.

[17] H. Ren and P. Kazanzides, Investigation of attitude tracking using an integrated inertial and magnetic navigation system for hand-held surgical instruments, *IEEE/ASME Trans. Mechatronics*, vol. 17, no. 2, pp. 210–217, Apr. 2012.

[18] S.Sabatelli, M.Galgani, L.Fanucci, A.Rocchi, "A Double-Stage Kalman Filter for Orientation Tracking With an Integrated Processor in 9-D IMU*", IEEE Trans. Instrum. Meas.*, vol. 62, No. 3, 590–598, Mar. 2013

[19] S. You, U. Neumann, and R. Azuma, Hybrid inertial and vision tracking for augmented reality registration. *Proc. of IEEE Virt. Real.*, Houston, TX, 260–267, 1999.

[20] S. You, U. Neumann, Fusion of vision and gyro tracking for robust augmented reality registration. *IEEE Virt. Real.,* Jan, 2001.

[21] N. Parnian, S.P. Won, F. Golnaraghi, Position Sensing Using Integration of a Vision System and Inertial Sensors. *34th Annu. Conf. IEEE Indust. Elec. Soc.*, pp 3011–3015, 2008.

[22] B. Hartmann, N. Link, G.F. Trommer, Indoor 3D Position Estimation Using Low-Cost Inertial Sensors and Marker-Based Video-Tracking, *IEEE/ION Pos. Loc. Navig. Symp.*, pp. 319–326, 2010.

[23] F. Raab, E. Blood, O. Steiner, and H. Jones, Magnetic position and orientation tracking system, *IEEE Trans. Aerosp. Electron. Syst.,* vol. AES-15, no. 5, pp. 709–717, May 1977.

[24] A. J. Chung, P. J. Edwards, F. Deligianni, and G. Z. Yang, Free-hand cocalibration of optical and electromagnetic trackers for navigated bronchoscopy, *Lect. Notes Comput. Sci.* Volume 3150, pp 320-328, 2004.

[25] A. Vaccarella, E. De Momi, A. Enquobahrie, G. Ferrigno, Unscented Kalman Filter Based Sensor Fusion for Robust Optical and Electromagnetic Tracking in Surgical Navigation, *IEEE Trans. Instru. Measur.* Volume: 62, pp 2067 – 2081, 2013.

[26] A. Tobergte, M. Pomarlan, G. Hirzinger, Robust Multi Sensor Pose Estimation for Medical Applications, *IEEE/RSJ Intern. Conf. on Intell. Robot. Sys.* , pp. 492-497, 2009.

[27] A. Tobergte, M. Pomarlan, G. Passig, and G. Hirzinger, "An approach to ulta-tightly coupled data fusion for handheld input devices in robotic surgery," *2011 IEEE Int. Conf. Robot. Autom.*, pp. 2424–2430, May 2011.

[28] G. Claasen, P. Martin, F. Picard, Optical-Inertial Tracking System with High Bandwidth and Low Latency, *Rec. Advan. Robot. Autom*, Volume 480, pp 171-181, 2013.

[29] S.J. Julier, The Scaled Unscented Transformation, Proc Amer. Contr. Conf., vol. 6, pp. 4555-4559, May 2002.

[30] P. Lang and A. Pinz, Calibration of hybrid vision/inertial tracking systems, *Proc. 2nd InerVis Workshop Integr. Vis. Inertial Sens.,* pp.1 -6, 2005.

[31] E. Dam, M. Koch, M. Lillholm, Quaternions, Interpolation and Animation*, Technical Report DIKU-TR-98/5,* Department of Computer Science, Univ. Copenhagen, Denmark, 1998.

[32] X. Pennec, Computing the mean of geometric features Application to the mean rotation, *Institut National de Recherche en Informatique et en Automatique* (IN-RIA), Le Chesnay, France, 1998.

[33] D. De Lorenzo, E. De Momi, L. Conti, E.Votta, M. Riva, E. Fava, and G. Ferrigno, Intraoperative forces and moments analysis on patient head clamp during awake brain surgery. *Med. Bio. Eng. Comput.* 51(3), 331-341, 2013.