Xsens MVN MotionGrid: Drift-Free Human Motion Tracking Using Tightly Coupled Ultra-Wideband and Miniature Inertial Sensors

Giovanni Bellusci, Daniel Roetenberg, Fred Dijkstra, Henk Luinge, and Per Slycke

***Abstract*—To support real-time, drift-free human motion capture of multiple closely interacting actors, Xsens introduces a revolutionary product: MVN Motiongrid. This system is based on tightly coupled ultra-wideband radio based positioning and inertial motion capture, and it is available as an optional add-on to the Xsens MVN full body motion capture system. This combination is capable of eliminating the incremental drift inevitable in motion capture based on inertial sensors alone, without affecting the unique advantages of MVN, namely immunity to occlusions, cost effective large area capture, and flexibility of use. This paper describes the basic working principles, architecture, and performance of Xsens MVN MotionGrid.**

I. Introduction

THE USE of inertial sensors has become a common practice in human motion tracking applications [1]-[7]. Unique advantages of this approach, compared for example to optical tracking systems, include immunity to occlusions and marker swapping, cost effective extremely large area tracking capabilities, very accurate and smooth joint angles, and flexibility in use due to the absence of any installed infrastructure. In this way, this technology can be employed in practically any environmental condition. However, any inertial based system will accumulate position tracking errors over time and traversed distance. In the Xsens MVN system [8], biomechanical joint constraints based on human body models are used to eliminate the integration drift of each body segment in relation to the others. The detection of external points on the segment with the world (e.g. foot contact) is used to limit the integration error of the assembled body model in the global frame. Since all segments are statistically connected by the biomechanical model [9]-[11], the contact constraint will implicitly update the kinematics of all segments and thus reduce the uncertainty of the global position. However, some inertial position drift on the horizontal plane is still present, typically between 1% and 2% of the traversed distance.

The presence of drift in the estimated position will not always be a problem, depending on the specific application; for example, in single actor scenarios, the virtual environment might be adjusted to coincide with the actor's actions. However, in this way any *real-time* operation is precluded, strongly limiting the number of potential situations in which the technology can be used.

MVN MotionGrid is a product by Xsens Technologies B.V., P.O. Box 559, 7500 AN Enschede, the Netherlands, T: +31 (0)889736700, F: +31 (0)889736701; www.xsens.com, patents pending.

In several scenarios, simultaneous motion capture of *multiple* actors might be required. This is a particularly critical situation to accomplish with inertial motion capture alone if the actors need to interact with each other or with the same objects, since they will not experience the same drift. Also in this case, at least to a certain extent, the relative drift can be corrected during post-processing (e.g. by editing foot contacts). However, this involves manual post-editing, and again, real-time applications are precluded. Other situations which demand for drift-free motion capture include: interaction with props, integration with other technologies (e.g. GPS in outdoor scenarios), and virtual camera systems.

Following these motivations, the possibility of adding an aiding technology to the MVN system, capable of eliminating the horizontal positioning drift is required for several applications. However, the choice of the aiding technology should not inhibit the unique advantages of the MVN system. In fact, adding an aiding positioning system to MVN also means introducing installed hardware for the reference frame creation. However, the amount of hardware that needs to be installed hugely varies with the type of technology used. For example, systems based on acoustic (e.g. ultrasound) signals or magnetic tracking will only give a workable and cost effective solution for small volumes due to their limited range. Optical systems can scale to reasonably large volumes but require a lot of hardware to do so, and obviously have the significant drawback of introducing occlusion issues, strongly limiting the scenarios and applications for which the technology can be used.

The aiding technology which has been chosen for MVN is pulse-based Ultra-Wideband (UWB) radio [12]-[15]. This emerging technology has unique capabilities and advantages compared to other positioning systems and perfectly fits the properties of MVN:

- UWB signals do not necessarily require line-of-sight and therefore are inherently much more robust to occlusion than optical systems; in fact, UWB radio waves can easily propagate through light obstructions like thin concrete, wood or glass walls. This possibility is precluded to systems based on vision.
- A much larger drift-free full body motion capture area can be constructed, compared to optical systems, with similar infrastructure. In fact, a standard UWB system can easily cover up to 20 m $\times$ 20 m, whereas a corresponding optical system would only be able to cover a motion capture area of a few meters (e.g. 3 m $\times$ 3 m).

- An UWB system can easily scale to *extremely large* volumes (e.g. a whole building) with only a fraction of the costs and installed infrastructure compared to optical systems.
- Due to the relatively limited amount of required infrastructure, the system is very easy to set-up and re-locate, it is portable.
- UWB does not suffer from restrictions in lighting or other environmental conditions (e.g. air pressure, moisture, temperature). This makes the system particularly flexible and suitable to be used in "non-laboratory" environments.
- Compared to other radio systems (e.g. GPS, WiFi), UWB has the unique advantage of providing centimeter level accuracy even in multipath indoor environments, thanks to the huge signal bandwidth used, typically in excess of a GHz. This unique characteristic is unmatched in any other wireless technology and makes UWB immune to reflections coming from e.g. metallic infrastructure present in the environment.

Additionally, a significant benefit of using UWB for position aiding is to make MVN much less dependent on magnetometers; in this way, the system is particularly robust to magnetic disturbance. This Xsens product, based on tight coupling [16] of inertial motion capture and UWB radio, is available as an optional add-on to the MVN system and it is called MVN MotionGrid ®. This paper describes the basic working principles, architecture and performance of this system, and it is organized as follows. In Section II, the generic concepts of UWB based position estimation are shortly introduced. Section III presents the MVN MotionGrid hardware, architecture, and system setup. In Section IV, the basic MVN MotionGrid working principles are described. Section V discusses the overall system performance based on an extensive data-set of measurements, and compares the achieved tracking accuracy with a traditional optical system. Finally, in Section VI, the main conclusions are drawn.

II. GENERIC PRINCIPLES OF UWB POSITIONING

UWB radio is an emerging wireless technology whose distinctive feature is the capability of transmitting very low-power and short duration (at nanosecond level) pulses with several GHz of bandwidth [12]-[14]. This unique characteristic provides this technology with a very high spatial resolution, which makes UWB ideal for centimeter level positioning applications based on *Time Of Arrival* (TOA) estimation [15], even in challenging multipath environments (e.g. inside buildings)[1]. For this reason, in the following, the generic principles of TOA based UWB positioning are introduced.

[1] In fact, the capability of resolving multipath components is directly related to the signal bandwidth. For example, an UWB radio system with 2 GHz of bandwidth (B) provides a multipath resolvability capability of approximately c/B = 15 cm, being c the speed of light in air. In this way, the TOA of the *direct* propagation path from transmitter to receiver, which contains the useful information for positioning applications, can be accurately estimated. It is also evident that, since the bandwidth of the GPS signal for civilian applications is only approximately 2 MHz, this technology is not suitable for accurate indoor positioning applications (the provided multipath resolvability is of about 150 m). Similarly, a WiFi system with 20 MHz of bandwidth would provide a multipath resolvability capability of about 15 m.

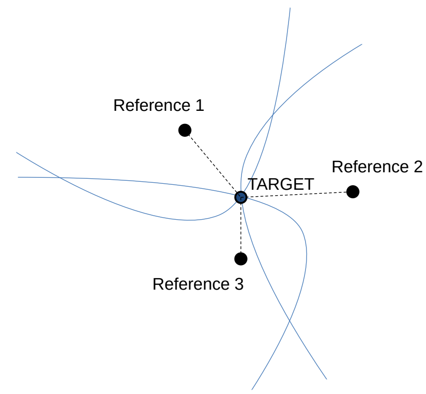

Fig. 1. Geometrical interpretation of the traditional approach to position estimation via TDOA based UWB ranging.

A. TOA based position estimation

TOA based position estimation relies on the measurement of the travel time of the radio signal propagating from a transmitter to a receiver. In the conceptually most simple situation in which two nodes have a common clock, the receiver node can determine the TOA of the incoming signal and directly calculate its distance from the transmitter, by multiplying the estimated TOA by the speed of light in air (after subtracting the known time of transmission). In this way, the estimated range allows to draw a circle (in 2-dimensions) with the reference node in its center and radius equal to the estimated range. By collecting at least three measurements between the target node with unknown position and some reference nodes with known position and by intersecting the defined circles, it is possible to determine the position of the target to be positioned. It should be noted that this very simple operation is made extremely critical by the stringent system timing requirements; in fact, timing errors of 3 nanoseconds would introduce distance errors of about 1 meter.

If there is no synchronization between the target and the reference nodes, but among the reference nodes only, the time of transmission is unknown as well. A common approach to solve this problem is to rely on Time Difference of Arrival (TDOA) techniques: the TDOA of two signals traveling between the target and two reference nodes is estimated; this determines an hyperbola, with foci at the two reference nodes, as shown in Fig. 1; the point of intersection of the hyperbolas provides the position of the target. This is the approach usually followed by UWB hardware manufactures; the drawback of this method is that the constructed TDOA measurements are no longer independently distributed. As an alternative, the time offset between the (synchronized) reference nodes and the target can be treated as an additional unknown in the non-linear system of positioning equations. As it will be further discussed in the next sections, this is the approach used in MVN MotionGrid.

III. MVN MOTIONGRID SYSTEM ARCHITECTURE AND SETUP

In this section, the specific hardware components and architecture of the MVN MotionGrid system are presented. Fig. 2 shows a schematic setup of the complete system.

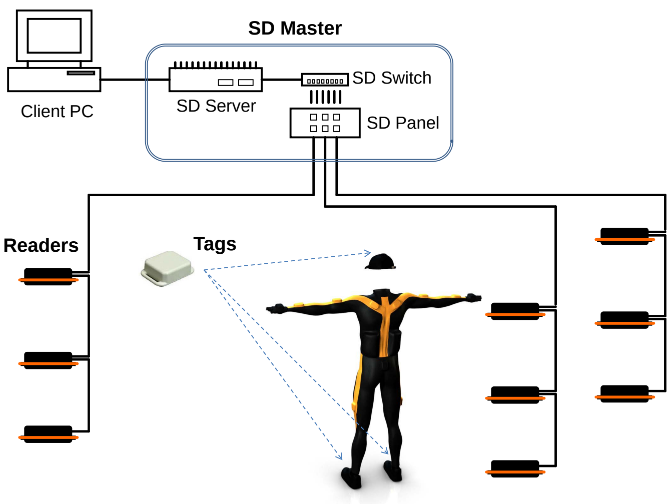

Fig. 2. Schematic representation of the complete Xsens MVN MotionGrid system.

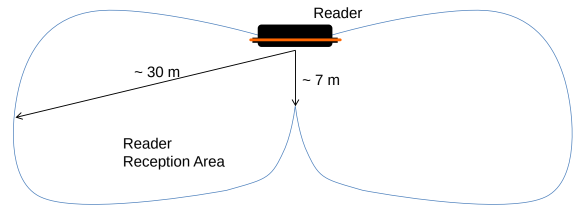

Fig. 3. Reader reception area.

Fig. 4. Standard 16 mm Reader mounting on the left of the figure; on the right, Reader attached to a tripod (above) and placed on a floor stand (below).

A. Tags

The Tags are the UWB radio transmitters of the MVN MotionGrid system. They are small (36 mm $\times$ 33 mm$\times$ 13 mm, weight 22 g), battery powered (battery life of about 1.2 years), and transmitting a burst RF-signal with -10 dB bandwidth of about 650 MHz, centered around 6.6 GHz. The extremely low power radiation (the average emitted power spectral density is lower than -41.3 dBm/MHz) makes this technology "invisible" to other wireless systems and allows it to coexist with other licensed radio devices without causing any harmful interference or performance impairment. With respect to the description in Section II, the Tags represent the target whose position needs to be determined. For optimal system performance, MVN MotionGrid uses three Tags placed on the top of the head and on each of the feet of the actor, as shown in Fig. 2; however, depending on the specific application requirements, other placements can be chosen.

B. Readers

The Readers are the UWB radio receivers of the MVN MotionGrid system. They demodulate the Tag data and measure the TOA for each Tag transmission. Each Reader is placed in a circular case (15.5 cm high, 35.5 cm diameter, total weight of 2 kg), which also contains an UWB antenna omni-directional in the horizontal plane. Fig. 3 illustrates the reception area in the vertical plane. As shown, the maximum coverage of a single Reader is of about 30 m, making it possible to use only a limited amount of infrastructure in order to cover very large drift-free motion capture volumes. The Readers can be easily installed on tripods, attached to walls or ceiling, or placed on the floor, using standard 16 mm mounting included in the system, as shown in Fig. 4.

With respect to the description in Section II, the Readers represent the network of reference nodes with known position, used to determine the Tag position.

C. Synchronization & Distribution Master

The Synchronization & Distribution (SD) Master is an intermediate processing device that serves as interface between the MVN MotionGrid Client (typically the PC on which MVN Studio is running) and the Readers. The SD Master, as shown in Fig. 2, is composed by three distinguished components:

- The SD Panel provides timing synchronization (at picosecond level) and power to the network of Readers. Up to three Readers can be daisy-chained to a single port of the SD Panel without requiring any external power supply, allowing in this way a great flexibility in use. The maximum length of the daisy-chain is defined by the Ethernet cabling specifications and is 90 m. If needed by the specific application, several SD Panels can be daisy-chained to allow for system scalability in case of particularly large setups.
- The SD Server is a fan-less PC that runs a Linux OS. It is the core of the SD Master and performs several configuration, administration and control tasks; moreover, it is responsible of converting the raw TOA packets received from the Readers into a single TOA sequence which can be used by the MVN MotionGrid Client.
- The SD Switch is a dedicated Ethernet switch which provides a communication link between the SD Server and the Readers.

The choice of dedicated hardware (the SD Server, physically distinguished from the Client PC) for processing the raw UWB data allows to control the delay introduced by the system and to meet the stringent performance requirements for real-time applications. In fact, the experienced overall latency is smaller than 20 ms, allowing a seamless integration of MVN MotionGrid with the MVN system.

D. MVN MotionGrid Readers constellations

The accuracy of the UWB position estimate is affected by two fundamental factors:

- The error statistics in each single range estimate between Reader and Tag. In the MVN MotionGrid system, in the absence of outliers, the pseudo-range error can be modeled as a zero mean Gaussian random variable with standard deviation of about 3 cm (intrinsic measurement noise).
- The particular choice of the number and location of the Readers in the capture volume, defined as *Readers constellation*. The effect of the Readers constellation can be described as a multiplicative factor, called Dilution Of Precision (DOP), which in each point of the capture volume indicates how the intrinsic measurement noise is amplified (or reduced) in the final UWB only position estimate. The DOP is therefore a purely geometrical effect which can be usefully employed for Readers constellation planning.

- Minimal and high-end constellations

As described in the previous subsections, the Readers are synchronized by the SD Panel, and therefore share a common clock. On the other hand, the clock of the Tags is not accurate enough to be able to predict at receiver side the time of transmission, and since only one-way communication is possible, it can not be synchronized to the Readers clock. Therefore, the time of transmission needs to be treated as an additional unknown to be determined together with the Tag coordinates; because of this, the minimum number of Readers needed to provide 3-D positioning is *four*. Larger number of Readers, however, will result in improved performance. This point is evident by comparing the two plots in Fig. 5, which shows a top view of the achievable *UWB only* 3-D positioning accuracy on an horizontal plane at 1.5 meters height, both for the minimum 4-Readers setup (above), and for a high-end setup with 12 Readers (below). In this last case, the achievable accuracy is better than 5 cm in all the area surrounded by the Readers constellation. Even more important, a high-end constellation provides significantly higher robustness towards outliers, which typically happen in UWB only positioning systems, guaranteeing centimeter level accuracy even when the Tag transmission to some Readers might be blocked or strongly attenuated due to the occurrence of Non Line-Of-Sight (NLOS). For this reason, Xsens recommends the use of 9 Readers for medium size setups with capture area up to 20 m × 20 m. Less Readers (e.g. 6), however, can be used for smaller size (5 m × 10 m) setups. It should be noted that, even with a high-end setup, the amount of infrastructure which needs to be added to MVN is still limited, resulting in a system particularly easy to set-up, move and reconfigure. This is also possible thanks to the very modest mounting requirements (e.g. the Readers are not sensitive to angular dependencies, like for examples optical cameras).

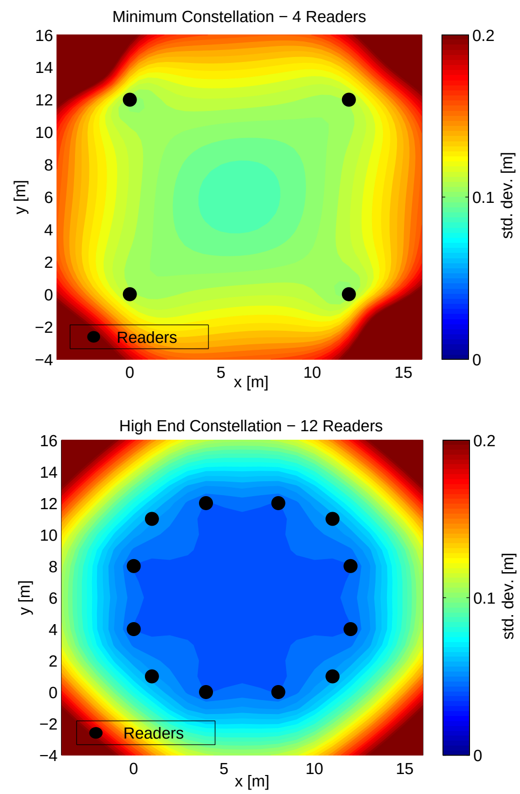

Fig. 5. Top view of achievable *UWB only* 3-D positioning accuracy on an horizontal plane at 1.5 meters height, both for the minimum 4-Readers setup (above), and for a high-end 12 Readers setup (below).

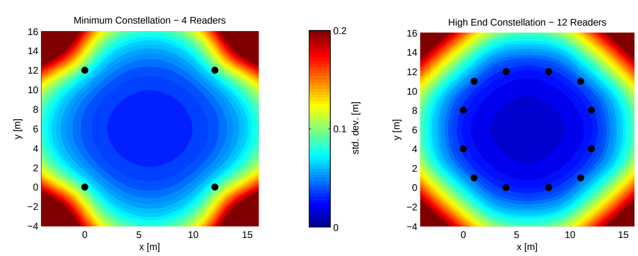

Fig. 6. Top view of achievable *UWB only* 2-D positioning accuracy on an horizontal plane at 1.5 meters height, for the same setups as in Fig. 5. In this case, the Tag height is known; this reflects the effective working principle of MVN MotionGrid, in which the Tag height is obtained from the MVN body model.

To guarantee superior achieved accuracy, in the current implementation of MVN MotionGrid, the height of each Tag is not calculated using UWB, but it is obtained from the MVN body model; therefore, UWB is used to provide the Tag position in the x-y plane only. For comparison, Fig. 6 shows the achievable *UWB only* 2-D positioning accuracy on the same horizontal plane as shown in Fig. 5; it is evident that, since in this case the Tag height is known, significantly better accuracy can be obtained.

- Freedom and flexibility in Readers constellation design

Due to the great flexibility in use, the Readers can be placed to satisfy specific application requirements. For example, in the left hand part of Fig. 7, the configuration could be used

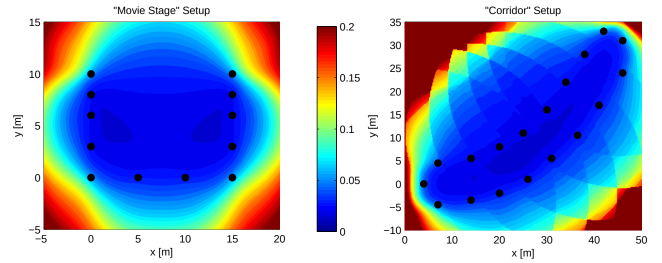

Fig. 7. Readers constellation for a movie setup (left) and for a 75 meters corridor shaped motion capture area (right).

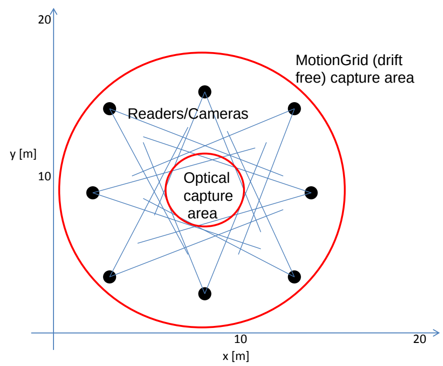

Fig. 8. Typical full body motion capture areas both for an optical system (approximately 3 m $\times$ 3 m), and for the MotionGrid system (approximately 15 m $\times$ 15 m), assuming the same position for the 8 optical cameras/Readers. The drift-free MotionGrid capture area is about 20 times larger, allowing an extremely efficient use of available space and significant saving on required square meters per capture volume.

on a movie stage. In this case, no Readers are placed on one side of the used volume to preserve an unobstructed view of the scene. Another example is shown on the right hand part of the same figure, where a 75 m "corridor" is created using 18 Readers. This might be necessary for example for outdoor sport applications.

- Total motion capture area vs. drift-free area

As can be seen from the previous figures, since the Readers have omni-directional antennas, the area in which UWB position is available and *drift-free* tracking is guaranteed extends beyond the area bounded by the Readers. This might be counter-intuitive to those experienced with optical systems in which the motion capture volume is typically significantly smaller than the volume bounded by the mounting points of the cameras. This point is shown in Fig. 8, in which the typical full body motion capture areas both for an optical system (approximately 3 m $\times$ 3 m), and for the MotionGrid system (approximately 15 m $\times$ 15 m) are given, assuming the same position for the 8 optical cameras/Readers considered in the example. The drift-free MotionGrid capture area is about 20 times larger, allowing an extremely efficient use of available space and significant saving on required square meters per capture volume.

It is also important to note that in all the examples given in the previous sections, the *actual motion capture* area is much larger than the area in which accurate drift correction can be

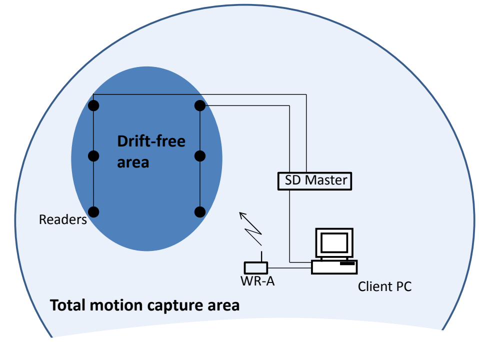

Fig. 9. Drift-free area vs. total motion capture area; the last one is only limited by the range of the MVN wireless receivers (WR-A) and it is typically in the order of a hundred of meters.

performed. In fact, the first is only limited by the range of the wireless receivers (WR-A), as illustrated in Fig. 9, which is typically in the order of a hundred of meters. The actor can freely exit or enter the drift free area, allowing a unique flexibility in use and performance.

E. MVN MotionGrid calibration

A preliminary operation which needs to be performed before using the MVN MotionGrid system, is represented by the system calibration. This operation is aimed at estimating the (initially) unknown position of the Readers and their relative clock-offset, and it is a prerequisite to achieve accurate positioning results. The calibration methods developed by the UWB hardware manufacturers only provide an estimate of the Readers clock-offset. The determination of the Readers position is left to the end-user and needs to be accurately manually surveyed. This operation is a time-costly and error-prone process, and only feasible for permanent setups with a limited number of Readers. Therefore it would negate the unique characteristics of MVN MotionGrid. For this reason, a novel easy-to-use calibration algorithm capable of accurately calibrating a UWB setup without additional hardware, and only requiring a couple of minutes and very simple user operations, has been developed and patented by Xsens. The following part of this section describes the basic principles of the automatic calibration.

Denoting with y_{mnk} the generic TOA measurement at Reader m for the $k-$th pulse from Tag n, results:

$$y_{mnk} = \tau_{nk} + \|\boldsymbol{r}_m - \boldsymbol{t}_{nk}\|_2 + \Delta\tau_m + \delta_{mnk} + e_{mnk} \quad (1)$$

where τ_{nk} is the time of transmission of the k-th pulse from Tag n, $\boldsymbol{t}_{nk}$ is the position of Tag n when transmitting the k-th pulse, and $\boldsymbol{r}_m$ and $\Delta\tau_m$ are the position and clock-offset of the m-th Reader. $\delta_{mnk} \geq 0$ is a possibly non-zero delay due to NLOS or multipath and e_{mnk} is the zero mean Gaussian distributed measurement noise with standard deviation σ. During the whole calibration, $\delta_{mnk} = 0$ is assumed; however, since this assumption will not always be valid, an outlier rejection algorithm is implemented. The description of this step goes beyond the scope of this paper and therefore it is omitted. The purpose of the calibration is to determine the Readers parameters $\boldsymbol{r}_m$ and $\Delta\tau_m$. The problem can be formulated as a constrained maximum likelihood estimation, yielding the following minimization:

$$\min_{\boldsymbol{\theta}} = \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{K_n}\epsilon_{mnk}^2(\theta) \tag{2}$$

$$\text{s.t. } A\boldsymbol{\theta} = \boldsymbol{b}$$

where:

$$\epsilon_{mnk}(\boldsymbol{\theta}) = \sigma^{-1}(\tau_{nk} + \|\boldsymbol{r}_m - \boldsymbol{t}_{nk}\|_2 + \Delta\tau_m - y_{mnk}) \tag{3}$$

are the normalized residuals, $\boldsymbol{\theta}$ is the parameters vector defined as:

$$\boldsymbol{\theta} = (\{\boldsymbol{t}_{nk}, \{\tau_{nk}\}_{k=1}^{K}\}_{n=1}^{N}, \{\boldsymbol{r}_m, \Delta\tau_m\}_{m=1}^{M}) \tag{4}$$

and M, N, and K_n represent the total number of Readers, Tags, and transmissions for the n-th Tag, respectively. The constraint in eq. (2), for proper choice of A and $\boldsymbol{b}$, only serves as a means of defining a reference system [18].

Since the minimization in eq. (2) is a non-convex problem, a reasonable initial estimate is needed to guarantee convergence to the correct solution. Therefore, before directly solving the optimization problem in eq. (2), an initial calibration is performed to provide a first approximate estimate of $\boldsymbol{\theta}$. This preliminary calibration consists of two steps. In the first step, initial estimates of the Readers position and clock-offset are determined. This operation uses the calibration Tags placed to the purpose inside the Readers case. In the second step, a 'Calibration Wand', containing a UWB Tag, is used; the user is asked to sway the Wand for about one minute, walking in all the area inside and around the Readers constellation. This simple and inexpensive operation is the only task the user needs to perform in order to calibrate the system. The estimates of the Readers positions and clock-offsets provided by the first step, are used in this second step, which gives as output an estimate of the Wand trajectory and the corresponding Tag time of transmissions. The overall initial estimates of the Readers positions and clock-offsets, and the positions and times of transmission of the Wand Tag, obtained as described in the two preliminary steps, are finally used as initial guess to solve eq. (2). It should be noted that the Readers position and clock-offset need to be determined together with the Wand Tag position and time of transmission. For this reason, the requirement for the minimum number of Readers to perform the calibration is *five*. More details on the principle of working and theoretical background of this solution can be found in [18].

The proposed calibration algorithm can provide a full 3-D Readers position estimate. However, to allow for improved calibration accuracy, in the MVN MotionGrid system the user is asked to provide the height of the Readers as input, since this is usually an inexpensive operation to perform. In fact, Readers placed on the ground have a known height and the height of the Readers attached to the ceiling can be easily measured with a laser distance tool included in the system. In this way, the major benefit of avoiding the costly and error prone surveying of the x and y Readers coordinates is preserved.

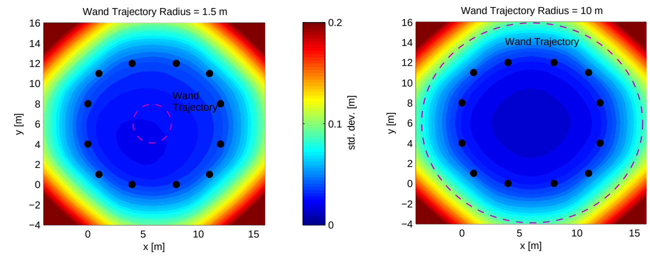

Fig. 10. Top view of the achievable UWB only 2-D positioning accuracy on an horizontal plane at 1.5 m height, for two different values of the calibration Wand Trajectory radius, for illustrative purposes.

Fig. 10 shows the top view of the achievable UWB only 2-D positioning accuracy on an horizontal plane at 1.5 m height, for the same high-end 12 Readers constellation shown in the right part of Fig. 6, for two different Wand trajectories, for illustrative purposes. The calibration accuracy is mainly affected by the area surveyed by the Wand; by surveying a sufficiently large volume (typically inside and around that bounded by the Readers), the achievable accuracy is practically the same as in case of perfect knowledge of the Readers position, as is clear from the plot on the right-end part of Fig. 10. This point reflects the unique characteristics of the Motiongrid system, in particular the large motion capture volume which extends beyond that bounded by the Readers mounting.

IV. MVN MotionGrid working principles

In this section, the basic working principles of the MVN MotionGrid system are presented. Referring to the general sensor fusion scheme described in [8] and reported in the following in Fig. 11 for convenience, the purpose of this section is to give insight on the approach used to seamlessly integrate UWB based position aiding in the MVN system This step is represented in the figure by the functional block: "Aiding sensors".

A. Tight coupling of inertial and UWB data

To guarantee superior robustness to occlusions, the current implementation of MVN MotionGrid makes use of three different UWB Tags, as show in Fig. 2. The Tag packet transmission frequency is 10 Hz and the Tags are not hardware synchronized to the corresponding inertial sensor; this provides significant advantages in terms of system flexibility. To guarantee that each Tag transmission is correctly aligned in time with the corresponding inertial sensor measurement data, synchronization algorithms are implemented in software. For the purpose of the discussion presented in this section, each of the three units composed by MT and Tag can be considered a single functional element. Therefore, in the following the coupling of the inertial and UWB data for a *single* MT plus Tag sensor unit, is presented. The way in which this information is used in the overall motion capture fusion engine will be shortly discussed in the next paragraph.

A *tightly* coupled sensor fusion approach, schematically represented on the left-hand part Fig. 12, is used to track

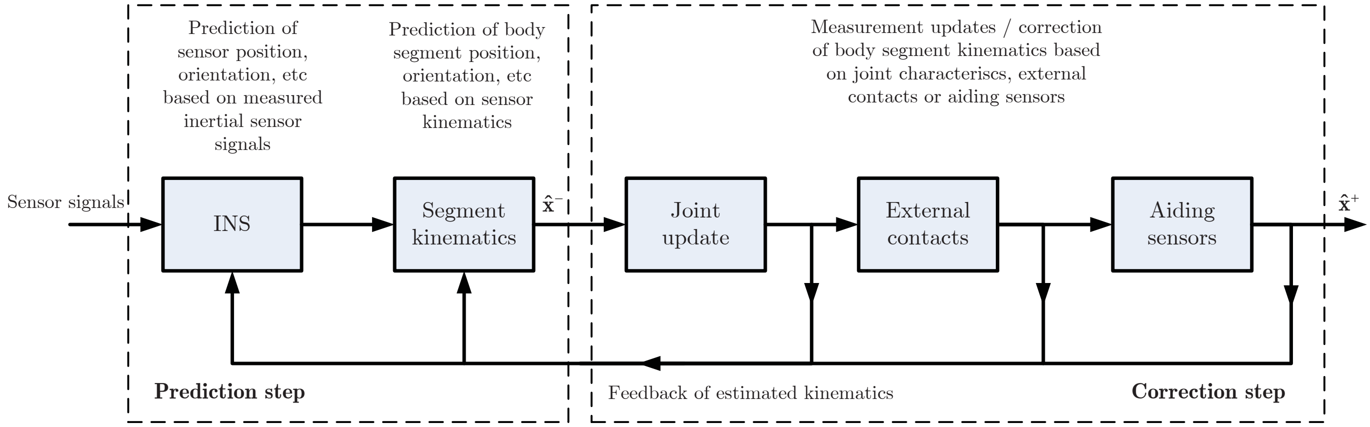

Fig. 11. MVN MotionGrid sensor fusion scheme as described in [8]. The coupling of the UWB measurements with the inertial data is represented by the functional block: ”Aiding sensors”.

the position and orientation of the MT plus Tag unit in a stationary coordinate frame, referred in the following as navigation frame. Also the UWB Readers position is expressed in this frame. In a tightly coupled approach, the ”raw” sensor measurements from the inertial sensing components (accelerometers, gyroscopes, and magnetometer) and the UWB TOA measurements are directly used for sensor fusion, instead of already filtered quantities like position or orientation, as shown on the right-hand part of the same figure for a loosely coupled sensor fusion approach. Hence, there is not an explicit multilateration step, as typically found in loosely coupled UWB positioning systems [19]-[20]; on the contrary, the multilateration of position is implicitly performed by the sensor fusion algorithm. The advantages of using this method are two-fold. Firstly, preprocessing of measurements typically results in loss of information. This is mainly due to approximations of statistical distributions, but in some cases measurements are ignored, for instance when there are not enough TOA measurements for multilateration. By directly using the sensor measurements, nothing has to be disregarded and maximal advantage is taken of the available information. Secondly, tightly coupled sensor fusion can perform hypothesis testing for the data belonging to the different sensors and efficiently deal with outliers. This is especially useful for UWB measurements, where outliers regularly occur due to multipath effects and/or NLOS conditions. Tightly coupled sensor fusion can disregard the affected measurements while still utilizing the remaining ones. Additionally, the available inertial information gives accurate predictions of the UWB measurements, which allows for improved outlier detection without the need to rely on motion models for the Tag (which would effectively work only for limited range of movement dynamics), as in the case of conventional UWB only positioning systems. Hence, a tightly coupled system is more robust.

Based on the system models and measurement noise of both the inertial and aiding system, an Extended Kalman Filter (EKF) gives an optimal estimate of the sensor kinematics [21]. The state-vector of the discrete-time non-linear state-space model can be expressed as:

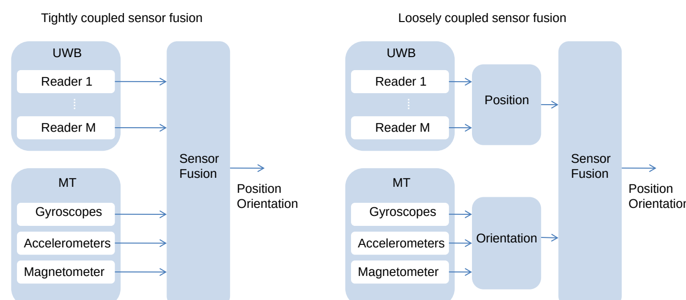

Fig. 12. On the left, tightly coupled sensor fusion used in MVN MotionGrid: the ”raw” measurements from the M UWB Readers and from the inertial sensor are directly used for sensor fusion. On the right, a traditional loosely coupled approach: in this case, already filtered quantities are used for sensor fusion.

$$\boldsymbol{x} = \left((\boldsymbol{b}^n)^T, (\dot{\boldsymbol{b}}^n)^T, (\boldsymbol{q}^{bn})^T, (\boldsymbol{\delta}^b_a)^T, (\boldsymbol{\delta}^b_\omega)^T, \tau, \dot{\tau} \right)^T \tag{5}$$

where $\boldsymbol{b}^n$, $\dot{\boldsymbol{b}}^n$ and $\boldsymbol{q}^{bn}$ denote the position, velocity and orientation quaternion of the sensor unit expressed in the navigation frame, and can be derived as shown in [22]. $\boldsymbol{\delta}^b_a$ and $\boldsymbol{\delta}^b_\omega$ are slowly time-varying inertial bias terms included in the process model as random walk, and τ is the Tag time of transmission modeled as an integrated random walk. The apex T denotes the sampling interval. The EKF handles the different sampling rates and a varying number of measurements straightforwardly. It runs at the high data rate of the inertial sensing unit and the 10 Hz UWB updates are only performed when measurements are available. Outliers from NLOS and/or multipath effects are detected using hypothesis testing on the residuals/innovations of the EKF. More details on the principle of working of the tightly coupled approach here presented can be found in [16].

B. *Position correction solver*

As shown in the sensor fusion scheme in Fig. 11, the aiding sensors are used to correct the position and orientation. Since the motion of the MVN character is constrained by a biomechanical model and floor contacts, the correction is not trivial. When there are no contacts (e.g. during running or jumping), the correction can be applied directly by using a weighted average of the estimated accuracy of the inertial integration and position accuracy of the aiding system. However, when there is a floor contact, the position of the complete body can generally not be corrected without introducing foot-slide. So, the most obvious solution is to correct the position difference by allowing foot-slide during floor contacts, which is available as a selectable scenario in MVN MotionGrid. However, in several motion capture applications, this kind of correction, especially the foot-slide, is not allowed. Therefore, in MVN MotionGrid, another scenario is offered to the user with a more sophisticated position correction solver. In this method, slight adjustments are made to the position and orientation of the segments based on the aiding position and inverse kinematics to gradually solve the position difference without introducing foot-slide. The time it takes to solve the gap depends on the amount of deviation, but it is implemented such that the overall pose of the character during solving is not altered noticeably.

V. MVN MotionGrid system performance

In this section, the performance of the MVN MotionGrid system is evaluated.

A. Comparison with an optical tracking system

In this subsection, the positioning accuracy provided by MVN MotionGrid is compared with an optical tracking system, used as reference for performance assessment.

The measurements have been collected in a 6 × 8 × 2.5 m room equipped with 8 optical cameras. The capture volume provided by the optical system is extremely limited, about 3 × 3 × 2 m. For this reason, only within this volume a performance comparison could be performed. An optical marker has been rigidly attached to the UWB Tag placed on the top of the head of the actor. The marker trajectory tracked by the optical system has been compared with the position provided by MVN in correspondence of the head point in which the Tag has been placed, both with and without using UWB for position aiding. For coordinate alignment between the two systems, a portion of the recorded measurements independent of that used for performance evaluation, has been used.

In the experiments hereby presented, a 10 Readers UWB constellation, placed on the border of the room, has been used. To give insight on the effective improvements provided by the MVN MotionGrid system, two different scenarios have been considered. In the first scenario, the actor is performing a regular walk of about 30 seconds in circles inside the optical capture volume. In the second one, after a 30 seconds walk inside the optical capture volume, used for coordinate systems alignment, the actor is leaving the volume and performing very fast movements, including significant foot sliding; this is expected to be a particularly critical situation for MVN only without any aiding technology. After this step, the actor is coming back to the optical capture volume and further walking for about 30 seconds. This part of the trial is used for performance evaluation.

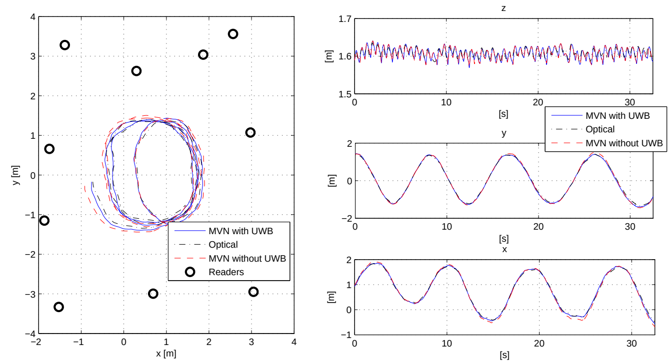

Fig. 13. Walking trajectories both for MVN (with and without UWB), and for the optical based tracking system, on the left-hand. On the right-hand, the performance comparison on each of the three axis, is shown.

Fig. 13 shows on the left-hand part a top view of the walked trajectory provided by MVN both with and without using UWB for aiding, for the first described scenario. It can be seen that in both cases the MVN trajectory is very close to that provided by the optical system. The very modest drift, present in the MVN only trajectory, is partly compensated by using UWB for position aiding; in fact, the 3-D root mean square (RMS) positioning error is in the two cases of about 0.065 m and 0.10 m, respectively. The right-hand part of the figure shows a comparison of the actor position on each of the three axis, separately, for the same three different cases. The higher accuracy on the z coordinate is due to the height aiding used in MVN (i.e. the known height of the head of the actor on which the Tag is mounted).

The achieved performance for the second scenario is shown in Fig. 14. On the left-hand part of it, the total walked trajectory, as recorded by MVN, is shown. As previously motivated, the comparison for this scenario has been done only in the final part of the walking, shown on the right-hand part of the same figure. As expected, without using any aiding technology, the positioning accuracy provided by MVN is significantly worsened and a clear misalignment with respect to the trajectory tracked by the optical system is visible; the 3-D RMS error is in this case of about 0.77 m. This is due to the very fast movements containing significant foot sliding performed by the actor, and therefore it should be intended as a worst case scenario to evaluate the positioning performance of the MVN only system. The use of UWB for aiding allows to significantly improve the positioning accuracy, even in this very challenging motion capture condition; in fact, the same figure also shows the trajectory tracked by MVN with UWB; in this case, the RMS error is of only about 0.10 m. The slightly higher value, compared to the corresponding one in the first scenario, is mainly due to a short initial transitory before converging to the correct position, while the actor is entering back in the volume in proximity of the Readers.

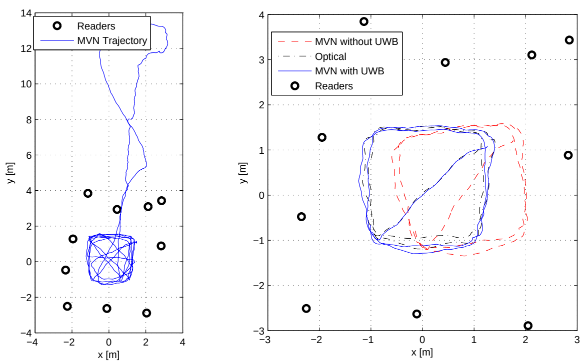

Fig. 14. Total walked trajectory, as recorded by MVN, for the second scenario, on the left. On the right, comparison between the walking trajectories provided both by MVN (with and without UWB), and by the optical tracking system.

B. MVN MotionGrid performance on a green screen virtual stage

In this subsection, the performance of MVN MotionGrid on a green screen virtual stage, is shown. Therefore, this section is representative of actual performance achievable with the system in a realistic user scenario. Fig. 15 shows the client virtual stage where the MVN MotionGrid system has been installed. A 12 Readers constellation has been employed to cover a drift-free motion capture volume of about $24 \times 10 \times 7$ m. Eight Readers have been placed on the ceiling, while four Readers on the ground. Note that one Reader on the ground has been placed behind the green plaster wall, proving a fair resistance of the UWB system towards occlusion issues. Due to the very large area covered, it was not possible to use an optical tracking system as reference for performance assessment. For this reason, the achieved accuracy has been evaluated by using a grid of 8 reference points spread in the volume, whose positions have accurately been surveyed. In order to give indication of the achievable accuracy in *all* the recording stage area, 6 of the 8 reference points have been placed in proximity of the border of the volume, where the DOP provided by the UWB Readers constellation is not optimal. The resulting 3-D RMS position error is in this case of about 7 cm. This value is fully in line with the experiments described in the previous section. On the contrary, typical errors achieved by the MVN only system, without the use of UWB position aiding, range between some decimeters and in excess of one meter, depending on the specific motion dynamics, and clearly prove the benefits provided by the MVN MotionGrid system in typical user scenarios of interest.

As illustrative example, Fig. 16 shows a screen-shot of a video of raw motion capture data as directly recorded on the green screen stage with the MVN Studio software. The full video can be found here. A single Tag placed on the head of the actor, as shown in the video, has been used in this case. The video shows the unique benefits of the MVN MotionGrid system: large motion capture volume, unobtrusive hardware setup, accurate and drift-free global position estimate, no occlusion or marker swapping issues; in short: Freedom of Movement.

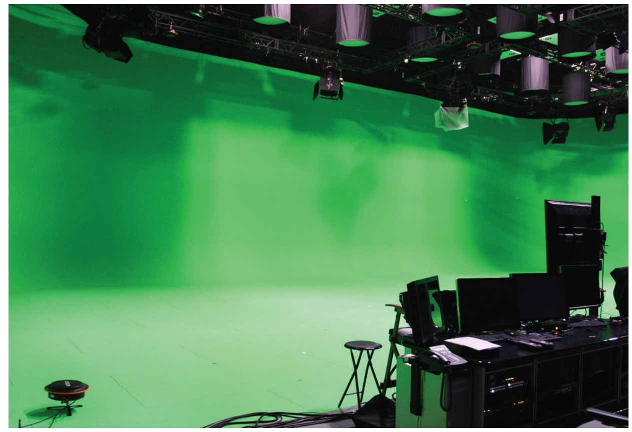

Fig. 15. MVN MotionGrid on a green screen virtual stage.

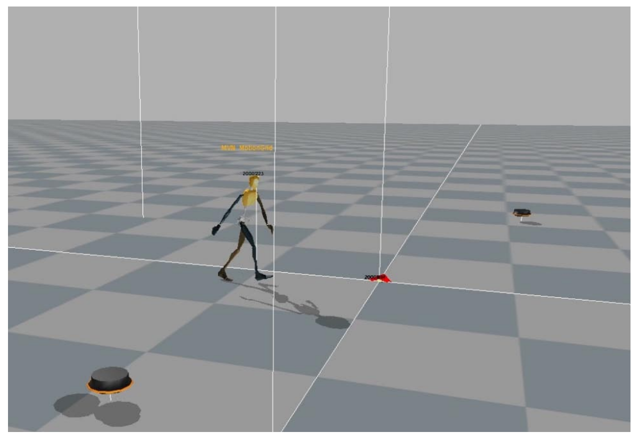

Fig. 16. Motion capture as directly recorded on a green screen virtual stage using the Xsens MVN Studio software; the full video is available here.

VI. Conclusion

In this paper, the basic working principles, architecture, and performance of the Xsens MVN MotionGrid, have been presented. This product is available as an optional add-on to the Xsens MVN system and it is based on tight coupling of inertial human motion capture and UWB radio based positioning. The choice of UWB radio as aiding technology allows to eliminate the incremental horizontal positioning drift inevitably present in inertial only based tracking, requiring only modest additional infrastructure and without affecting the unique advantages of the MVN system: robustness, large area capture volumes, and flexibility of use. Comparisons with an optical based tracking system show that MVN MotionGrid is able to provide centimeter level absolute positioning accuracy, even in critical motion capture conditions.

References

[1] J. Morris, *"Accelerometry - A technique for the measurement of human body movements"*, Journal of Biomechanics, vol. 6, pp. 729-736, 1973.
[2] D. Roetenberg, P. Slycke, P. Veltink, *"Ambulatory position and orientation tracking fusing magnetic and inertial sensing"*, IEEE Transactions on Biomedical Engineering, vol. 54, pp. 883-890, 2007.

[3] D. Roetenberg, *"Inertial and magnetic sensing of human motion",* PhD Thesis, University of Twente, 2006.
[4] P. Bonato, *"Wearable sensors/systems and their impact on biomedical engineering",* IEEE Engineering in Medicine and Biology Magazine, vol. 22, no. 3, pp. 18-20, 2003.
[5] E. Foxlin, *"Inertial head-tracker sensor fusion by a complementary separate-bias Kalman filter",* in Proceedings of VRAIS 96, 1996, pp. 185-194.
[6] E. Bachmann, *"Wearable sensors/systems and their impact on biomedical engineering",* PhD Thesis, Naval Postgraduate School, 2000.
[7] T. Molet, R. Boulic, and D. Thalmann, *"Human motion capture driven by orientation measurements",* Presence, vol. 8, no. 2, pp. 187-203, 1999.
[8] D. Roetenberg, H. Luinge, P. Slycke, *"Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors",* Xsens Technologies B.V. White Paper, Apr. 2009.
[9] R. Zhu and Z. Zhou, *"A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package",* IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 12, no. 2, pp. 295-302, 2004.
[10] V. Zatsiorsky, *"Kinematics of Human Motion. Champaign III: Human Kinetics",* 1998.
[11] D. Giansanti, V. Macellari, G. Maccioni, and V. Macellari, *"The development and test of a device for the reconstruction of 3-D position and orientation by means of a kinematic sensor assembly with rate gyroscopes and accelerometers",* IEEE Transactions on Biomedical Engineering, vol. 52, no. 7, pp. 1271-1277, 2005.
[12] M.Z. Win, R.A. Scholtz, *"Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications",* IEEE Transaction on Communications, vol. 48, no. 4, pp. 679-691, Apr. 2000.
[13] FCC, Office of Engineering and Technology, *"Revision of part 15 of the Commission's rules regarding ultra-wideband transmission systems",* ET Docket, no. 98-153, 2002.
[14] Liuqing Yang, G.B. Giannakis, *"Ultra-wideband communications: an idea whose time has come",* IEEE Signal Processing Magazine, vol. 21, no. 6, pp. 26-54, Nov. 2004.
[15] S. Gezici, Zhi Tian, G.B. Giannakis, H. Kobayashi, A.F. Molisch, H.V. Poor, Z. Sahinoglu, *"Localization via ultra-wideband radios: a look at positioning aspects for future sensor networks",* IEEE Signal Processing Magazine, vol. 22, no. 4, pp. 70-84, Jul. 2005.
[16] J.D. Hol, F. Dijkstra, H. Luinge, T.B. Schon, *"Tightly Coupled UWB/IMU pose estimation",* in Proceedings of IEEE ICUWB 2009, pp. 688-692, Sep. 2009.
[17] R. Roberts, *"Ranging subcommittee final report",* Nov. 2004.
[18] J.D. Hol, T.B. Schčn, F. Gustafsson, *"Ultra-wideband calibration for indoor positioning",* accepted for presentation, IEEE ICUWB 2010, Sep. 2010.
[19] S. Sczyslo, J. Schroeder, S. Galler, and T. Kaiser, *"Hybrid localization using UWB and inertial sensors",* in Proceedings of IEEE ICUWB 2008, pp. 89-92, Sep. 2008.
[20] S. Pittet, V. Renaudin, B. Merminod, and M. Kasser, *"UWB and MEMS based indoor navigation",* The Journal of Navigation, vol. 61, no. 3, pp. 369384, Jul. 2008.
[21] T. Kailath, A. H. Sayed, and B. Hassibi, *"Linear Estimation",* Prentice-Hall, Inc, 2000.
[22] J.D. Hol, *"Pose estimation and calibration algorithms for vision and inertial sensors",* Lic. Thesis No 1379, Dept. Electr. Engineering, Linkčping University, Sweden, May 2008.