Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors
Daniel Roetenberg, Henk Luinge, and Per Slycke

Abstract—The Xsens MVN motion capture suit is an easy-to-use, cost efficient system for full-body human motion capture. MVN is based on unique, state-of-the-art miniature inertial sensors, biomechanical models and sensor fusion algorithms. MVN does not need external cameras, emitters or markers. It can thus be used outdoors as well as indoors, there are no restrictions for lighting, it does not suffer from problems of occlusion or missing markers. In addition, unique for inertial motion capture technology: the sensor-suit captures any type of movement, including running, jumping, crawling and cartwheels.

I. INTRODUCTION

To capture human body motion in an ambulatory situation without the need for external emitters or cameras, several systems are available. Mechanical trackers utilize rigid or flexible goniometers which are worn by the user. These angle measuring devices provide joint angle data to kinematic algorithms which are used to determine body posture. Attachment of the body-based linkages as well as the positioning of the goniometers present several problems. The soft tissue of the body allows the position of the linkages relative to the body to change as motion occurs. Even without these changes, alignment of the goniometer with body joints is difficult, especially for multiple degree of freedom joints.

The use of inertial sensors has become a common practice in ambulatory motion analysis [1, 2]. For accurate and drift free orientation estimation several methods have been reported combining the signals from 3D gyroscopes, accelerometers and magnetometers [3, 4]. Accelerometers are used to determine the direction of the local vertical by sensing acceleration due to gravity. Magnetic sensors provide stability in the horizontal plane by sensing the direction of the earth magnetic field like a compass. Data from these complementary sensors can be used to eliminate drift by continuous correction of the orientation obtained by integrating rate sensor data.

By using the calculated orientations of individual body segments and the knowledge about the segment lengths, rotations between segments can be estimated and a position of the segments can be derived under strict assumptions of a linked kinematic chain [4–6]. This method assumes an articulated rigid body in which the joints only have rotational degrees of freedom. However, a human body and its joints cannot be modeled as a pure kinematic chain with well-defined joints such as hinge-joints and ball-and-socket-joints. Each human joint allows some laxity in all directions (both position and orientation) other than its main direction of movement [7]. Moreover, to be able to track complex human joints and non-rigid body parts such as the back and shoulder accurately, more than three degrees of freedom, as given by an orientation measurement, are required. Furthermore, importantly, with only orientation driven motion capture, it is not possible to analyze the clearance of both feet, which occurs during running or jumping. Using this approach, it is also not possible to accurately determine the displacement of the body with respect to a coordinate system not fixed to the body.

To provide full six-degree-of-freedom tracking of body segments with connected inertial sensor modules, each body segment’s orientation and position can be estimated by, respectively, integrating the gyroscope data and double integrating the accelerometer data in time. However, due to the inherent integration drift, these uncorrected estimates are only accurate within a few seconds [8]. By combining the inertial estimates with other body worn aiding systems, such as an acoustic [9] or a magnetic tracker [10], unbound integration drift can be prevented.

Fig. 1. Xsens MVN consists of 17 inertial and magnetic sensor modules. Data is transmitted by a wireless connection to the laptop computer on which the processing is performed and visualized. A suit is used for quick and convenient placement of sensors and cables.
The MVN motion capture system is fully ambulatory and consists only of body worn sensors. The system is unique in its approach to estimate body segment orientation and position changes by integration of gyroscope and accelerometer signals which are continuously updated by using a biomechanical model of the human body. This allows for tracking of dynamic motion. By facilitating the constraints of the model, notably, the segments are connected by joints, the kinematics of body segments are corrected for drift and other errors. The system runs in real-time with a maximum update rate for all kinematics of 120 Hz. With the MVN Studio software, the user can easily observe, record and export the movements in 3D. This paper describes the design, working principles, output formats and technical features of MVN.

II. WORKING PRINCIPLES

A. System set-up

The MVN motion capture system consists of 17 MTx sensors with two Xbus Masters [11]. The MTx is an inertial and magnetic measurement unit and comprises 3D gyroscopes, 3D accelerometers and 3D magnetometers (38×53×21 mm, 30 g). The sensor modules are daisy chained connected to the Xbus Masters, meaning that there is only one cable leading to each limb. The Xbus Masters synchronizes all sensor sampling, provides sensors with power and handles the wireless communication with the PC or laptop. For quick and convenient placement, the sensors and cables are integrated in a Lycra suit and the Xbus Masters are mounted on the back. The total weight of the system (including 8 AA batteries) is 1.9 kg. Sensor modules are placed on the feet, lower legs, upper legs, pelvis, shoulders, sternum, head, upper arms, fore arms and hands (see Figure 1).

When attaching sensors to a body, the initial pose between the sensors and body segments is unknown. Moreover, the assessment of distances between body segments is difficult by numerical integration of acceleration because of the unknown initial position. Therefore, a calibration procedure has to be performed in which the sensor to body alignment and body dimensions are determined (Section II-B).

Since the sensor signals and the biomechanical model can be described in a stochastic manner, it can be incorporated in a sensor fusion scheme with a prediction and correction step (see Figure 2). In the prediction step, all sensor signals are processed using so-called inertial navigation system (INS) algorithms (Section II-C). This is followed by the prediction of the segment kinematics using a known sensor to body alignment and a biomechanical model of the body (Section II-D). Over time, integration of inertial sensor data leads to drift errors due to presence of sensor noise, sensor signal offset, or sensor orientation errors. To correct the estimated quantities, such as orientation, velocity, and position, the sensor fusion scheme updates the estimates continuously. The correction step includes updates based on biomechanical characteristics of the human body (Section II-E), notably joints, detection of contact points of the body with an external world which constrains the global position and velocity (Section II-F), and, optionally, other aiding sensors. Estimated kinematics are fed back to the INS algorithms and segment kinematic step to be used in the next time frame.

B. System calibration

To express segment kinematics in the global frame, the kinematics of the sensors must be subjected to a step of calibration wherein the orientation of the sensor module with respect to the segment (see Figure 4b) and the relative distances between joints are determined.

To find the sensor to segment alignment, several methods are used and combined. In the first step, the subject is asked to stand in an a priori known pose: the T-pose (upright with arms horizontally and thumbs forward). The rotation from sensor to body segment \( B_S \mathbf{q} \) is determined by matching the orientation of the sensor in the global frame \( G_S \mathbf{q} \) with the known orientation of each segment \( G_B \mathbf{q} \) in this pose.

\[
G_B \mathbf{q} = G_S \mathbf{q} \otimes B_S \mathbf{q}^* \quad (1)
\]

where \( \otimes \) denotes a quaternion multiplication and \( * \) the complex conjugate of the quaternion [12].
In the second step, the subject is asked to perform a certain movement that is assumed to correspond to a certain axis. For example, the arm axis is defined by a pronation or supination movement. The measured orientation and angular velocity are used to find the sensor orientation with respect to the segment’s functional axes [13].

Initial estimates of joints positions are obtained by measuring several body dimensions: body height, arm span and foot size. For subject specific scaling, these estimates can be refined by measuring several anatomical landmarks. With each additional provided dimension, the scaling model is adjusted. Dimensions which can be entered include the greater trochanter, lateral epicondyle on the femoral bone, lateral malleolus, anterior sup. ilias spine and the acromion. Other dimensions are obtained by using regression equations based on anthropometric models [14–16].

As a final step in the calibration procedure, the sensor to segment alignment and segments lengths can be re-estimated by using a priori knowledge about the distance between two points in a kinematic chain. For example, when the subject holds his hands together while moving them around, the distance between the left and the right hand palm is zero for each pose (see Figure 3). This closed kinematic chain can be solved which will improve the calibration values.

![Fig. 3. Hand touch calibration: the sensor to segment alignment and segment lengths are re-estimated by solving the closed kinematic chain for each pose.](image)

### C. Inertial tracking

Rate gyroscopes measure angular velocity $\omega$, and if integrated over time, provide the change in angle (or orientation) with respect to an initially known angle:

$$G^{S}q_{t} = \frac{1}{2}G^{S}q_{t} \otimes \Omega_{t}$$

where $G^{S}q_{t}$ is the quaternion describing the rotation from sensor $(S)$ to global frame $(G)$ at time $t$. $\Omega_{t} = (0, \omega_{x}, \omega_{y}, \omega_{z})^{T}$ is the quaternion representation of the angular velocity $\omega_{t}$.

Linear accelerometers measure the vector of acceleration $a$ and gravitational acceleration $g$ in sensor coordinates. The sensor signals can be expressed in the global reference system if the orientation of the sensor $G^{S}q_{t}$ is known:

$$G^{S}a_{t} = G^{S}q_{t} \otimes (S^{S}a_{t} - S^{S}g) \otimes G^{S}q_{t}^{*}$$

After removing the gravity component, the acceleration $a_{t}$ can be integrated once to velocity $v_{t}$ and twice to position $p_{t}$, all in the global frame:

$$G^{S}p_{t} = G^{S}a_{t}$$

![Fig. 4. Calibration and segment kinematics step.](image)

### D. Segment kinematics

The INS kinematics are translated to body segment kinematics using a biomechanical model which assumes that a subject’s body includes body segments linked by joints and that the sensors are attached to the subject’s body segments. Joint origins are determined by the anatomical frame and are defined in the center of the functional axes with the directions of the X, Y and Z being related to functional movements (see Figure 4a). For example, flexion/extension of the knee is described by the rotation about the $BY$-axis of the lower leg with respect to the upper leg; abduction/adduction is the rotation about the $BX$-axis; and endo/exo rotation is about the $BZ$-axis.

When the position $p_{U0}$ of the joint origin, the orientation $GBq_{U}$, and the length $s_{U}$ of segment $U$ are known, the position of point $p_{U1}$ in the global frame can be calculated (assuming a rigid segment) (see Figure 4c):

$$G^{S}p_{U1} = G^{S}p_{U0} + GBq_{U} \otimes B_{S}p_{U0} \otimes GBq_{U}^{*}$$

At $t = 0$, the origin of segment $L$, point $p_{L0}$, is connected to point $p_{U1}$ of segment $U$ (see Figure 4d).

The biomechanical model consists of 23 segments: pelvis, L5, L3, T12, T8, neck, head, and right and left shoulder, upper arm, fore arm, hand, upper leg, lower leg, foot and toe. For the segments on which no sensor is attached, the kinematics are estimated based on the biomechanical model incorporating stiffness parameters between connecting segments.

Skin and soft tissue artifacts do influence the measurements of the sensors, similar to other skin-based systems due to
contracting muscles (active) and skin and fat (passive) [17]. As a result, position or orientation changes in the segments around a joint can be measured which are biomechanically unlikely. In the joint measurement update step, these artifacts are reduced by using the knowledge that two segments are on average connected but with a statistical uncertainty. For specific joints, rotational characteristics are also described in statistical terms. For example, the knee is modeled as a soft hinge: the main axis of rotation is flexion and extension whereas internal rotation and abduction are limited to a few degrees and thus statistically more unlikely.

E. Joint update

After each inertial and segment kinematic prediction step, the uncertainty of the joint position and rotation will grow due to sensor noise and movement related errors (e.g. soft tissue artifacts) (see Figure 5a). These will be corrected using the joint measurement updates.

Within the scope of this paper, it is not feasible to provide detail of each kinematic update. We will focus on the positional constraint of a joint and its formulation in a Kalman filter. For each joint, the position relation can be expressed as a linearized function:

\[ y_t = Cx_t + w_t \]  

where \( x_t \) is the state vector at time \( t \) containing the positions of the two segments \( U \) and \( L \), \( C \) is the measurement matrix relating the state vector to the measurement \( y_t \), \( w_t \) is the measurement noise. When two segments are connected, measurement matrix \( C \) is given by:

\[ C = \begin{bmatrix} I_3 & -I_3 \end{bmatrix} \]  

\( I_3 \) symbolizes the 3 by 3 identity matrix.

A Kalman filter is used to estimate the state using the joint relation and the state prediction by the segment kinematic integration step:

\[ \hat{x}^+_t = \hat{x}^-_t + K(y_t - C\hat{x}^-_t) \]

where \( \hat{x}^-_t \) and \( \hat{x}^+_t \) are the states before and after the Kalman update, respectively, and \( K \) is the Kalman gain [18]. The Kalman gain is computed based on stochastic parameters about positional and rotational characteristics for each joint and propagation of errors by each integration step based on the sensor noise. With the Kalman filter update, the kinematics are corrected for drift and the uncertainty of the joint position is reduced (see Figure 5b).

To correct orientation errors due to gyro integration errors, the gravity vector as measured with the accelerometers provides inclination stability as described in [19]. To correct rotation errors about the vertical, a magnetic sensor is used as a heading aiding sensor. The earth magnetic field is locally easily disturbed by metallic objects, for example by structural beams in buildings or in automobiles, which can influence the measurements. In sensor fusion scheme, this disturbance is also estimated each time step as part of the process, similar to the approach in [20]. This results in a high degree of immunity against distortions. Moreover, a real-time visual warning is given in case of magnetic disturbances.

F. Contact detection and Aiding sensors

The assumptions about joints in an articulated body are used to eliminate the integration drift of each segment in relation to each other, while the detection of external points on the segment with the world is used to limit the boundless integration error of the assembled body model in the global frame. Under most circumstances, it can be assumed that the body must be in contact with an external physical world and is subject to gravity.

The probability of external contacts is based on computed kinematics, such as the position, velocity and acceleration, of relevant body landmarks. External contacts with the physical world are not limited to feet touching the ground but can also occur at other parts of the human body such as hands, knees, etc. In the default scenario, it is assumed that the floor is a flat surface. This means that in the applied contact constraint, the vertical position is updated being zero. When an external contact is detected, the position and velocity estimates of this point are updated using the sensor fusion scheme algorithms as described above. Because all segments are statistically connected by the biomechanical model, the contact constraint will implicitly update the kinematics of all segments and thus reduce the uncertainty of the global position. Other scenarios.
are also available such as ‘pelvis fixed’, which can be useful in situations where the person is sitting and moving around (e.g. on a horse or bike). Moreover, contacts can be set/overruled at a specified height (or 3D position) which enables tracking in irregular terrain such as jumping from a table or swinging on a high bar.

Without aiding, the initial starting position of the assembled body in the horizontal plane is not related to an external reference system and is therefore set at zero. The accuracy of the horizontal translation of the system is about 2% of the travelled distance. To improve the accuracy of captured motion data or to reference to an external coordinate system, for example in estimating the position of the entire body in space, the sensor fusion scheme can seamlessly be integrated with various types of aiding sensors, such as GPS, RF-based local positioning sensors, a barometer, a camera, as well as pressure and/or force sensors. For each aiding sensor, statistical information about its accuracy is necessary. In [21], position information of reflective markers as used in optical mocap was combined with inertial sensors.

III. OUTPUT - SEGMENT KINEMATICS

The sensor fusion scheme calculates the position, velocity, acceleration, orientation, angular velocity and angular acceleration of each body segment, $B$, with respect to an global (earth-fixed) reference coordinate system, $G$ (see Figure 7). By default, the global reference coordinate system used is defined as a right handed Cartesian coordinate system with:

- X positive when pointing to the local magnetic North.
- Y according to right handed coordinates (West).
- Z positive when pointing up.

Note that in some studies, different anatomical and reference frames definitions are used.

A. Joint angles

Typically, a joint rotation is defined as the orientation of a distal segment $^{G_B}q_L$ with respect to a proximal segment $^{G_B}q_U$: 

$$^{B}q_{UL} = ^{G_B}q_U \otimes ^{G_B}q_L$$  \hspace{1cm} (9)

There are a few commonly used parameterizations of the joint rotation $^{B}q_{UL}$ that describe joint angles:

- the Cardan/Euler representation [22, 23]
- joint coordinate system [24]
- helical angle [25]

All representations of the joint angle are based on the same quaternion or rotation matrix; the differences lies only in how the angles are extracted from this rotation. The International Society of Biomechanics (ISB) has proposed standards for rotations sequences for the lower [26] and upper body [27]. Note that positions of anatomical landmarks are not measured directly as in optical mocap systems, but computed using the measured segment kinematics in combination with the anatomical model. However, in many clinical protocols, joint angles are evaluated, which are directly measured with the sensors [28].

IV. OUTPUT - ANIMATION FORMATS

Recorded mocap data can be easily be exported to formats such as BVH [29] or FBX, which can be imported directly in 3D applications such as 3ds Max, XSI, MotionBuilder and others (see Figure 6). MVN Studio also features streaming output to MotionBuilder [30] which allows real-time character animation, including retargeting.
A. Retargeting

In character animation, the dimensions and joints of the character (creature) often do not exactly match the subject (actor) being captured. This can result in conflicting marker and joint angle data between the character and subject. The problem of adapting a recorded motion of a subject to a character is called retargeting. Many animation software packages have various tools for forward/inverse kinematics solving, retargeting, etc. An example of MVN data imported in MotionBuilder can be seen in Figure 8.

![MVN output of subject playing basketball](a)

![Retargeting of mocap data to a character](b)

Fig. 8. (a) MVN output of subject playing basketball. (b) Retargeting of mocap data to a character.

V. MVN STUDIO

MVN Studio is an easy-to-use graphical user interface and allows the user to observe the movements of the subject in real-time and from previous recordings (Figure 9). The data can be stored in the native formats MVN, which conserves all raw sensor data and can be used for (non-casual) re-processing, and MVNX, the MVN Open XML format (text/ASCII based) which contains all kinematic data. Mocap data can be exported to standard animation formats (BVH and FBX). MVN Studio is equipped with the option to capture up to four actors at the same time. The recordings are synchronized and can be viewed and edited in one viewport. The MVN Studio SDK is able to process real-time 3D position aiding input (i.e. optical or GPS-based).

The MVN motion capture system can be used anywhere, in or outside within the range of the wireless link. There is no need to install or place any fixed infrastructure. The total set-up time (including calibration) takes less than 10 minutes. MVN is a lightweight on-body system and can be worn under normal clothing or for example a green suit. There is no limitation by occlusion with either objects surrounding or other persons interacting with the subject. Subjects are not forced to a specific measurement volume and their movements can be measured in a familiar environment while performing their tasks - as in daily life.

![Screenshot MVN Studio](image)

Fig. 9. Screenshot MVN Studio.

REFERENCES


