# Motion Capture: from Radio Signals to Inertial Signals

Matteo Giuberti and Gianluigi Ferrari

**Abstract** The study of the motion of individuals allows to gather relevant information on a person status, to be used in several fields (e.g., medical, sport, and entertainment). Over the past decade, the research activity in motion capture has benefited from the progress of portable and mobile sensors, paving the way toward the use of motion capture techniques in mHealth applications (e.g., remote monitoring of patients, and telerehabilitation). Indeed, even if the optical motion capture, which typically relies on a set of fixed cameras and body-worn reflecting markers, is generally perceived as the standard reference approach, other motion capture techniques, such as radio and inertial, are attracting an increasing attention because of their suitability in remote mHealth applications.

Moreover, several hybrid approaches have been studied and proposed in order to overcome the limitations of component technologies considered independently. In this chapter, we present an overview of possible integration strategies between radio and inertial motion capture techniques. We start by investigating a radio-based approach, based on the fingerprinting radio localization technique. Then, the previous approach is improved by integrating inertial measurements: namely, accelerometers are used to provide an estimate of the nodes' pitches. Finally, the radio signals are abandoned in favor of only inertial measurements (obtained through accelerometers, gyroscopes, and magnetometers). The advantages and limitations of all approaches are discussed in a comparative way, characterizing the similarities and differences between the various approaches.

---

Matteo Giuberti*
Xsens Technologies B.V.
Enschede, The Netherlands
e-mail: matteo.giuberti@xsens.com
*MG was with University of Parma while performing this work.

Gianluigi Ferrari
Department of Information Engineering
University of Parma
Parma, Italy
e-mail: gianluigi.ferrari@unipr.it

# 1 Introduction

The health of an individual is mostly assessed through the evaluation of specific biophysiological signals. In particular, the evolution over time of such signals may give the doctors (or caregivers) very relevant information about the evolution of the pathology itself. However, this typically requires a frequent, or sometimes constant, monitoring of the patient and, thus, a large use of human resources.

Parameters such as blood-pressure, heart-pulse-rate, and respiration are common examples of this class of biophysiological signals. Besides them, it has been shown that the motion of individuals allows to gather relevant information on a person status, especially in many healthcare scenarios characterized by patients affected by motion disorders (e.g., Parkinson's Disease or Stroke) or, more simply, by patients who need a constant control of their movements (e.g., post-rehabilitation after surgery or monitoring of activities of elderly people).

Furthermore, over the past decade, the progress of portable and mobile sensors has paved the way toward the use of these sensors in mHealth applications, allowing a remote evaluation of different biophysiological signals with economic savings (in terms of specialized human resources) and improved patient comfort.

Even if optical motion capture, which often relies on a set of fixed cameras and body-worn reflecting markers, is generally perceived as the standard reference approach for the evaluation of the motion of individuals, other motion capture techniques, such as radio and inertial, are attracting an increasing attention because of their suitability in remote mHealth applications. Moreover, several hybrid approaches have been studied and proposed in order to overcome the limitations of component technologies considered independently. More specifically, in the literature, two main approaches have been mainly considered for motion capture: optical and inertial.

Concerning *optical motion capture*, the widely used technology is optoelectronic (e.g., Vicon system [3]). Optoelectronic systems require the user to wear reflective markers and to move in a space completely visible by a set of cameras. Because of their accuracy, these systems are often used as ground truth references for other motion capture systems. On the other hand, their use is mostly limited to clinical environments or specialized laboratories, due to their high cost and complexity. Other optical motion capture systems comprise the class of markerless systems. Microsoft Kinect [2], which makes use of an RGB camera and a depth sensor (composed of an infrared camera and projector), can be considered one of the most significant examples of markerless systems. Its low cost (with respect to systems like Vicon), along with its still quite good performance, has made it a widely used solution. Typically, optical systems suffer from problems related to different lighting conditions and markers' occlusion. Moreover, the user movements must be limited to the area captured by the cameras. The reader is referred to [8, 22] for accurate surveys on optical motion capture.

Concerning *inertial motion capture*, inertial sensors are used to estimate the orientation of rigid body segments and, thus, to capture the motion of a user. One of the most successful and complete commercial products is the Xsens MVN [27],

which comprises 17 inertial sensors (equipped with triaxial accelerometers, gyroscopes, and magnetometers) attached to the body of the user by a Lycra suit. The major advantage of this technology, with respect to optical systems, is that the user is completely free to move everywhere because no camera is needed. Moreover, the visibility of the nodes placed on the user body is not an issue. However, the accuracy of these systems is typically lower than that of optical systems and the cost, particularly for systems which rely on a large number of nodes and types of sensors (as Xsens MVN), is not significantly lower than that of optoelectronic systems. Finally, especially when used for a long time, a significant drift in the sensors' measurements can be observed, leading to a performance degradation.

When the motion of the whole body needs not to be tracked, besides complex systems like Xsens MVN, simpler configurations of nodes can be used. To this end, many other subsystems have been studied and developed in order to provide simpler solutions to simpler scenarios. These systems still rely on the use of accelerometers, gyroscopes, magnetometers, and/or combinations of them for different types of biomedical applications [7, 18, 29, 38, 42].

Finally, solutions based on the joint use of different technologies, designed in order to tackle and overcome the limitations which characterize component technologies considered independently, have also been investigated. For instance, inertial/radio, inertial/GPS, inertial/optical, and inertial/acoustic joint measurements are considered in [11, 12, 21], in [6], in [32], and in [36], respectively. In particular: in the inertial/radio approach, radio signal strengths are used to determine (after a preliminary training phase) a first rough estimate of the user's pose, eventually corrected through inertial measurements. In the inertial/GPS approach, nodes equipped with GPS are also used to track the user, providing better performance (especially) in outdoor scenarios (e.g., sport sessions, such as skiing). In the inertial/optical approach, inertial sensor nodes and cameras are jointly used, leading to a decreased user's freedom of movement but also to an improved robustness to occlusions. In the inertial/acoustic approach, nodes equipped with microphones and speakers are used in addition to inertial sensor nodes, providing better performance (thanks to the estimates of relative distances between nodes) but suffering from "acoustic occlusions."

Since motion capture applications typically rely on the use of a centralized base station, which collects all the data generated from the sensor networks, they are very suitable to be used in mHealth scenarios where, for instance, a mobile phone can directly communicate with the base station to get the processed data and send them to some specialized personnel. This is also reflected by various efforts already spent in developing mHealth applications based, for instance, on the use of inertial sensors [5, 23, 24, 25, 37, 41].

This chapter aims at presenting a comparative overview of the possible integration between radio and inertial approaches, with specific focus on the motion capture of an arm. In Sect. 2, the considered approaches are properly introduced and described; the mathematical background behind each approach is also briefly presented along with some performance results. In Sect. 3, a comparative overview between the selected approaches is given: to this end, the advantages and limita-

tions of every approach are discussed, trying to motivate the transitions between different approaches and to highlight their relationships. In Sect. 4, two application case studies are presented. In Sect. 5, future directions of the presented techniques and technologies are discussed. Finally, conclusions are drawn in Sect. 6.

## 2 Motion Capture: from Radio Signals to Inertial Signals

In this section, different approaches to (arm) motion capture are described. The mathematical background behind each approach is briefly presented and some performance results are also discussed. In particular, we start by investigating a radio-based approach (Sect. 2.1), based on the fingerprinting radio localization technique. Then, the previous approach is improved by integrating inertial measurements (Sect. 2.2): namely, accelerometers are used to provide an estimate of the nodes' pitches. Finally, the radio signals are abandoned in favor of only inertial measurements obtained through accelerometers, gyroscopes, and magnetometers (Sect. 2.3).

### *2.1 The Radio Fingerprinting-based Approach*

Fingerprinting is a robust localization technique for indoor scenarios, which are typically characterized by reflections, multipath, and fading [34]. We now provide some intuition on the fingerprinting technique—the interested user can find more details in [11, 12, 21]. Fingerprinting requires three kinds of nodes: target nodes, anchor nodes, and a base station. Target nodes have to be localized, whereas anchors nodes have fixed known positions and are used to generate a reference system. Finally, the base station is the processing center. Two phases are involved: a training phase, during which a radio map of fingerprints is generated; and an online phase, during which localization is performed.

   During the training phase, the target node continuously broadcasts packets to be received by the anchor nodes. The latter, upon reception of the packets sent by the target, measure the Received radio Signal Strengths (RSSs) and relay this information to the base station. The base station collects the RSS values and generates a "fingerprint," i.e., a vector containing the average RSSs measured by the anchors. Different positions of the target (i.e., fingerprint positions) in the monitored area (properly chosen depending on the considered application) correspond to different fingerprints. The set of fingerprints created in the training phase represents the "radio map" of the environment. The major strength of fingerprinting consists of the fact that the fingerprint vectors implicitly take into account the impact of reflections and multipath on the RSSs, i.e., it is "tailored" to the specific indoor environment. This makes fingerprinting virtually insensitive to indoor propagation limitations—provided that the propagation environment remains quasi-static. The radio map (ob-

tained from the average RSSs) can be used to run either deterministic (based on simple comparisons between the RSSs and the fingerprints) or probabilistic (based on a more accurate statistical characterization of the RSS) localization algorithms. To this end, during the training phase the entire Probability Mass Functions (PMFs) of the RSS from each anchor need to be also computed and stored.

After the training phase is completed, the online phase starts (following the same operations of the training phase, but now building a so-called "online" vector) and the radio map (deterministic approach) or the PMFs (probabilistic approach) created in the training phase are used to localize the target. Note that, at this stage, the target node can move freely and should not necessarily be placed in the fingerprint positions. Given the measured "online" vectors, different algorithms can be used to estimate the target positions.

One of the simplest *deterministic* fingerprinting algorithm is the Nearest Neighbor (NN) algorithm, whose generalization is known as *k*NN [9]. It consists of the computation of a specific distance metric between the received online vector and every fingerprint. According to the *k*NN algorithm, by applying the Shepard method [30], the estimated target position $\hat{\boldsymbol{s}} = (x, y, z)$ is given by

$$\widehat{\boldsymbol{s}} = \sum_{i=1}^{k} \frac{w_i}{\sum_{j=1}^{k} w_j} \cdot \widehat{\boldsymbol{s}}_i \qquad (1)$$

where $\{\widehat{\boldsymbol{s}}_i\}_{i=1}^{k}$ are the positions of the $k$ best neighbors (i.e., those with shortest distances from the target) and

$$w_i \triangleq \frac{1}{d_i^{p_s} + 0.0001} \qquad (2)$$

where $d_i$ is the distance computed between the $i$-th neighbor and the target (defined in the space of RSS vectors) and $p_s$ is an integer larger than 0. The term 0.0001 is used to prevent a division by zero if the online vector is equal to one of the fingerprints. Various definitions of distance can be considered, such as Euclidean and Manhattan [9].[1] Observe that, when $k = 1$ (i.e., with the NN algorithm), Eq. 1 reduces to the coordinates of the closest fingerprint position and, then, $p_s$ has no influence on the system. Finally, note that, due to the interpolation between fingerprint positions in Eq. 1, the estimated position $\widehat{\boldsymbol{s}}$ may (likely) differ from any of the considered fingerprint position.

Unlike the deterministic approach, in the *probabilistic* approach (straightforwardly called p-*k*NN) the RSSs measured at the anchor nodes are characterized, using the samples received in the training phase, through their entire PMFs. More precisely, the estimated target position can be expressed as

---

[1] Other distance definitions can be applied, e.g., the Mahalanobis distance, which takes also into account the contribution of covariance matrix computed for every fingerprint [9]. Little performance differences are, however, observed.

$$\widehat{\boldsymbol{s}} = \sum_{i=1}^{k} \frac{P\left(\widehat{\boldsymbol{s}}_i | \boldsymbol{r}\right)}{\sum_{j=1}^{k} P\left(\widehat{\boldsymbol{s}}_j | \boldsymbol{r}\right)} \widehat{\boldsymbol{s}}_i \tag{3}$$

where: $\boldsymbol{r}$ is the "online" vector; $P\left(\widehat{\boldsymbol{s}}_i | \boldsymbol{r}\right)$ is the a-posteriori probability of the $i$-th (out of $k$) best neighbor; and the $k$ best neighbors are chosen so that the corresponding a-posteriori probability is maximized. Note again that, when $k = 1$ (and thus reducing to a so-called p-NN), Eq. 3 returns exactly the coordinates of the closest fingerprint position.

Even if in common practice fingerprinting is often used to localize subjects (with accuracy in the order of a few meters) in large (indoor) areas [4, 13, 31], it can be also exploited to estimate the positions of (i.e., localize) multiple target nodes, properly placed on a user arm (e.g., one on the upper arm and one on the forearm), and straightforwardly derive the whole arm posture (and motion). To this end, a fixed origin must be properly chosen (e.g., the shoulder) and the user must try to keep this origin fixed during the evaluation. In such scenario, the fingerprint positions are pre-defined postures of the arm that must be held by the user during the *training* phase. During the following *online* phase, the user can instead move his/her arm freely (always recalling not to move the reference system origin, i.e., his/her shoulder).

In [11, 21], a detailed experimental analysis was carried out for the motion capture of the arm of an individual. Specifically, 3 SunSPOTs, acting as targets, were placed on the right arm of the user: the first node was on the hand, the second on the forearm, and the third on the upper arm. In the surroundings of the user, $N = 8$ anchor nodes were considered. An illustrative representation of the overall experimental set-up is shown in Fig. 1. Four postures of the arm (on the basis of a priori known movements of the arm) were trained leading to the computation of 12 fingerprint positions (4 positions per target node) and the origin of the reference system was fixed on the shoulder. In the online phase, the system performance was tested asking the user both to replicate trained postures and to perform different (intermediate) ones.
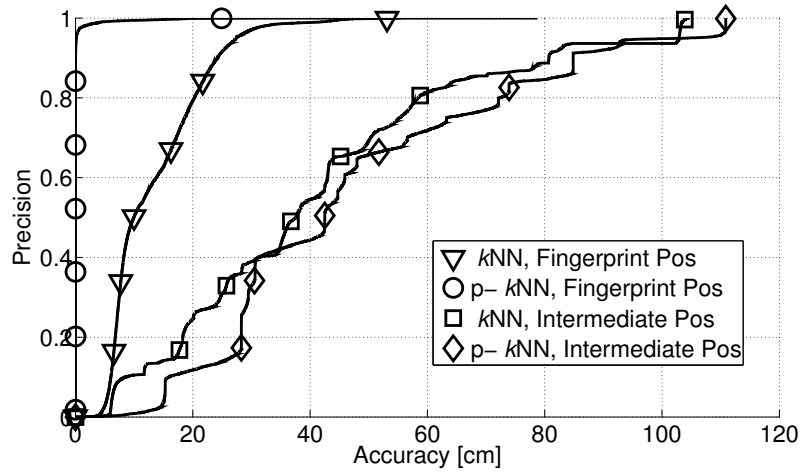
In Fig. 2, the precision is shown, as a function of the accuracy, for both the $k$NN and the p-$k$NN algorithms.[2] Either fingerprint or intermediate positions are considered during the online phase. The results show that the system performance, when estimating intermediate positions, is remarkably worse than that in correspondence to the estimation of the fingerprint positions. Note also that the $k$NN algorithm performs better than the p-$k$NN algorithm when the target is positioned in the intermediate positions. On the other hand, when the target is in one of the fingerprint positions, the p-$k$NN algorithm has a better performance. This is due to the fact that the probabilistic approach tends to give more weight to the fingerprint positions.

---

[2] Note that the performance graph can be read as follows: percentage of correct position estimates (i.e., precision) as a function of a given maximum tolerable error (i.e., accuracy).
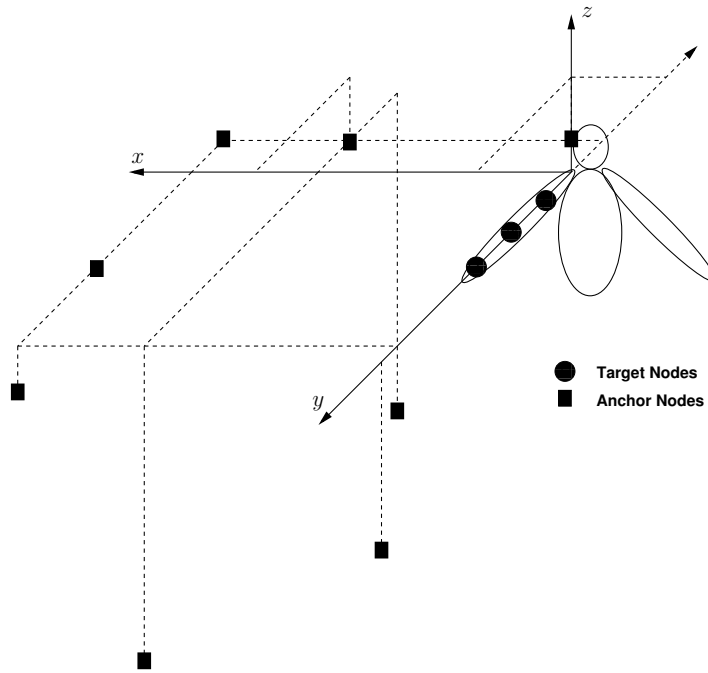
**Fig. 1** Considered motion capture set-up: placement of 3 SunSPOTs (i.e., target nodes) on the user arm and overall scenario.



**Fig. 2** Precision, as a function of the accuracy, for both the $k$NN and the p-$k$NN algorithms (using the optimized values $p_s = 1$ and $k = 2$). The system performance is shown for the estimation of both fingerprint and intermediate positions.

## 2.2 The Radio/Accelerometric Approach

As previously observed, the radio fingerprinting approach described in Sect. 2.1 may introduce errors, especially when a target node is in a position that differs from the trained fingerprint positions (i.e., when the arm is in an untrained posture). A possible way to improve the system performance is to estimate the arm orientation by making use of other inertial sensors (e.g., accelerometers, gyroscopes, and/or magnetometers), which the target nodes can be equipped with. In particular, considering proper combinations of these sensors (e.g., an accelerometer and a gyroscope), the orientation of a device (and, thus, of the arm) can be estimated [20].

Generally, the orientation of a device can be described by three parameters: yaw (or heading), pitch (or elevation), and roll (or bank) [16]. In order to define them, one must consider two reference systems: (i) the device coordinate system (that can be considered equal to the coordinate system of the inertial sensor mounted on it) and (ii) the Earth coordinate system. It is known that a rigid body can be arbitrarily rotated by first rotating it around its $z$ axis by an angle $\psi$ (the yaw), then around its $y$ axis by an angle $\theta$ (the pitch), and finally around its $x$ axis by an angle $\phi$ (the roll) [16]. It can be shown that, using just an accelerometer (in order to minimize the system cost), the pitch of a *still* device can be estimated as follows [40]:

$$\theta = \arcsin \overline{a}_{\mathrm{arm}} \qquad (4)$$

where $\overline{a}_{\mathrm{arm}}$ is the normalized acceleration (in g units) measured along the arm axis (which is the device axis that points toward the direction of the user arm, as shown in Fig. 3). Note that, according to Eq. 4, the pitch belongs to $[-\pi/2, \pi/2]$.

As the above approach is valid only if the device is still, static (as the constant force of gravity) and dynamic (as the movements or vibrations of the accelerometer itself) accelerations need to be discriminated. This problem is generally solved by applying a low-pass filter to the acceleration outputs or by taking into account only acceleration measurements with amplitudes in $[g - \xi, g + \xi]$, where $g$ is the gravity acceleration (i.e., 9.81 m/s$^2$) and $\xi$ needs to be properly chosen with respect to the application context (i.e., considering only data segments in which the user is not moving and, thus, $\xi \cong 0$ m/s$^2$).

Furthermore, if a device is placed on a user body, its position can be estimated more accurately by taking into account the implicit constraints of the human body. The following assumptions are expedient to derive useful human body constraints:

- the pitch of each device is known (e.g., using Eq. 4);
- the lengths of the body segments on which the devices are mounted (e.g., the length of the arm between two joints) are known;
- preliminary estimates of the positions of the devices are available (e.g., from radio localization).

Focusing on the motion capture of a user arm, assume that two nodes are placed on the user arm: the first on the upper arm and the second on the forearm. For ease of clarity, let us denote their physical positions as $\{\mathbf{D}_i\}_{i=1}^2$, where $\mathbf{D}_i = (x_{\mathrm{D}_i}, y_{\mathrm{D}_i}, z_{\mathrm{D}_i})$.
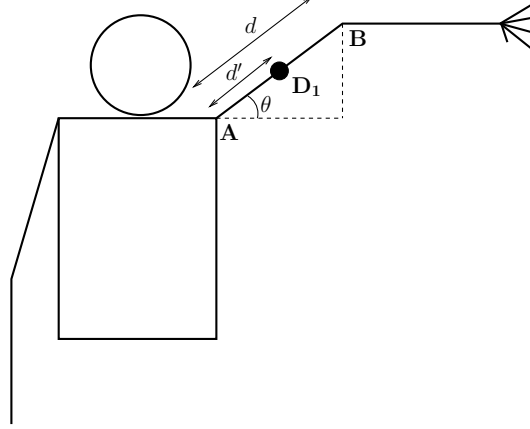
**Fig. 3** Arm constraints: key points and distances used in the algorithm, focusing on the first step of the proposed approach. In particular: **A** and **B** are the three-dimensional coordinates of the joints of the arm segment considered in the first step of the proposed approach (i.e., shoulder and elbow); $\mathbf{D}_1$ indicates the three-dimensional coordinates of target 1; $d$ is the actual length of the considered arm segment (in this case, the upper arm); and $d'$ is the distance between a joint and the next (starting from the shoulder) sensor device (in this case, the distance between the shoulder and target 1).

The origin $(0,0,0)$ will be the shoulder. Another reference point will be the joint of the arm between the upper arm and the forearm, namely the elbow.

First, the "reference pitch" of each device has to be defined. In our case, we define it as the pitch that every device has when the user raises his/her arm in a direction perpendicular to the vertical axis of his/her body. As shown in Fig. 3, at the first step the arm reference point, corresponding to the shoulder, is denoted as point $\mathbf{A} = (x_A, y_A, z_A)$ and the nearest joint (i.e., the elbow) as point $\mathbf{B} = (x_B, y_B, z_B)$. We then consider the first device, denoted as point $\mathbf{D}_1$ and positioned in the middle of the upper arm, i.e., between $\mathbf{A}$ and $\mathbf{B}$. Its initial position estimate (namely, from radio localization) is denoted as $\widehat{\mathbf{D}}_1$ and will be refined using the pitch information (recovered through Eq. 4). In order to do this, we define the two values $d'$ and $d$ as the actual distances between $\mathbf{A}$ and $\mathbf{D}_1$ and between $\mathbf{A}$ and $\mathbf{B}$, respectively. Knowing the device pitch $\theta$, it can be shown (through geometric considerations relative to the scenario in Fig. 3) that the adjusted position estimate of the first device, denoted as $\overline{\mathbf{D}}_1$, can be expressed as

$$\overline{\mathbf{D}}_1^{\mathrm{T}} = \begin{pmatrix} x_{\overline{D}_1} \\ y_{\overline{D}_1} \\ z_{\overline{D}_1} \end{pmatrix} = \begin{pmatrix} x_A + R\left(x_{\widehat{D}_1} - x_A\right) \\ y_A + R\left(y_{\widehat{D}_1} - y_A\right) \\ z_A + d'\sin\theta \end{pmatrix} \tag{5}$$

where $(\cdot)^{\mathrm{T}}$ denotes vector transposition and

$$R = \frac{d' \cos\theta}{\sqrt{\left(x_A - x_{\widehat{D}_1}\right)^2 + \left(y_A - y_{\widehat{D}_1}\right)^2}}. \tag{6}$$

On the basis of the estimate $\overline{\mathbf{D}}_1$ and taking into account the coordinates of $\mathbf{A}$ and the pitch $\theta$, the coordinates of $\mathbf{B}$ (namely, the elbow) can be expressed as follows:

$$\mathbf{B}^T = \begin{pmatrix} x_A + K\left(x_{\overline{D}_1} - x_A\right) \\ y_A + K\left(y_{\overline{D}_1} - y_A\right) \\ z_A + d\sin\theta \end{pmatrix} \tag{7}$$

where $K \triangleq d/d'$.

The above estimation procedure can be repeated to estimate the position of the second device $\mathbf{D}_2$, positioned in the middle of the forearm, between point $\mathbf{B}$ (the elbow) and point $\mathbf{C}$ (the wrist), as shown in Fig. 3. To this end, Eqs. 5 - 6 are used again considering proper values of $\theta$, $d'$, and $d$ (namely, the ones related to the arm segment between $\mathbf{B}$ and $\mathbf{C}$), and substituting $\mathbf{A}$ and $\mathbf{D}_1$ with $\mathbf{B}$ and $\mathbf{D}_2$, respectively.[3] Note that as the position of $\mathbf{B}$ is estimated (i.e., $\mathbf{B}$ is not a reference point as $\mathbf{A}$), it is likely that the estimate of the coordinates of $\mathbf{D}_2$ will be less accurate than that of the coordinates of $\mathbf{D}_1$.

According to the described hybrid approach, the $z$ coordinate is estimated relying only on the pitch estimation (which depends on the accelerometric measurements), whereas $x$ and $y$ coordinates are first coarsely estimated through the radio localization and then their estimates are refined (more precisely, rescaled) according to the pitch estimation. Alternatively, we can interpret our motion capture system as a system able to estimate the user arm heading (i.e., the direction of the arm in a horizontal plane) and pitch (i.e., the inclination of the arm). According to this point of view, the roles of the two component techniques in our hybrid approach (i.e., radio localization and accelerometric measurements) are even more definite. Indeed, it can be observed that the arm pitch estimation is carried out exclusively by using the accelerometric measurements, whereas the arm heading is estimated only through radio localization.

Finally, we also remark that the integration of the accelerometers into the system does not have any impact on the training phase. Indeed, all the acceleration measurements are taken into account only during the online phase.

In [12], this hybrid motion capture system was then tested for a scenario similar to the one described in Sect. 2.1. In particular, a body area network with target nodes on the user arm was still considered, but now the anchor nodes were in part placed (and fixed) in the surroundings of the user and in part attached on his/her body. This was obtained through the use of (i) a home-made t-shirt with folders where anchors can be placed and (ii) a hat with an anchor attached to it. Though the presence of some anchor nodes at fixed (outside the body) positions still forces the user to re-

---

[3] Observe that this process can be iteratively repeated, should more consecutive body segments be considered (e.g., by adding a device for the hand in our testbed).
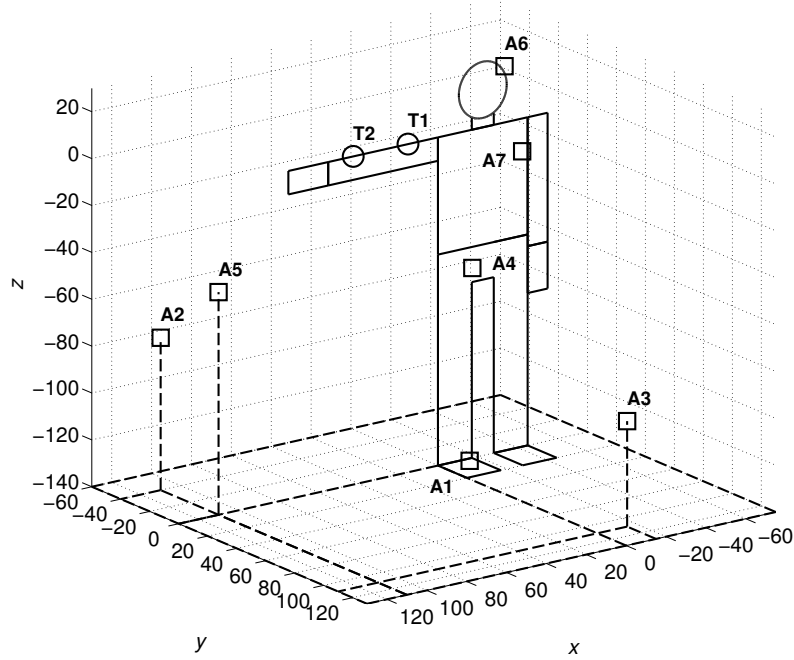
**Fig. 4** Experimental motion capture set-up for the radio/accelerometric approach. Note that A2, A3, and A5 are fixed in the surroundings of the user, whereas A1, A4, A6, and A7 are attached on his/her body.

main in its initial position in the room (in order to make the fingerprinting technique work consistently), the considered testbed was a first step toward a fully portable arm motion capture system. Furthermore, a noisier (and, thus, more realistic) environment was there taken into account (e.g., with more than one people moving in the room while testing the system), whereas in [11, 21] the system performance was analyzed in more controlled (interference-free) scenarios.

In the experimental testbed, only 2 SunSPOTs, acting as targets (i.e., target 1 and target 2), were placed on the right arm of the user: the first node (i.e., target 1) is on the upper arm and the second (i.e., target 2) on the forearm. $N = 7$ anchor nodes were considered: 4 of them were placed on the user body, whereas the remaining 3 were placed in its proximity. For ease of clarity, an illustrative representation of the overall experimental set-up is shown in Fig. 4, where a map of the positions of the nodes is shown. Still fixing the origin of reference system on the shoulder (and thus forcing the user to move arm and keep the shoulder still), 5 fingerprint positions per target node were considered, corresponding to 5 simple arm postures. The choice of the 5 fingerprint positions was made in order to cover as uniformly as possible the surroundings of the user arm.

During the online phase, the user was first asked to replicate the 5 training fingerprint positions (in order to test the system in *static* conditions). Furthermore, a
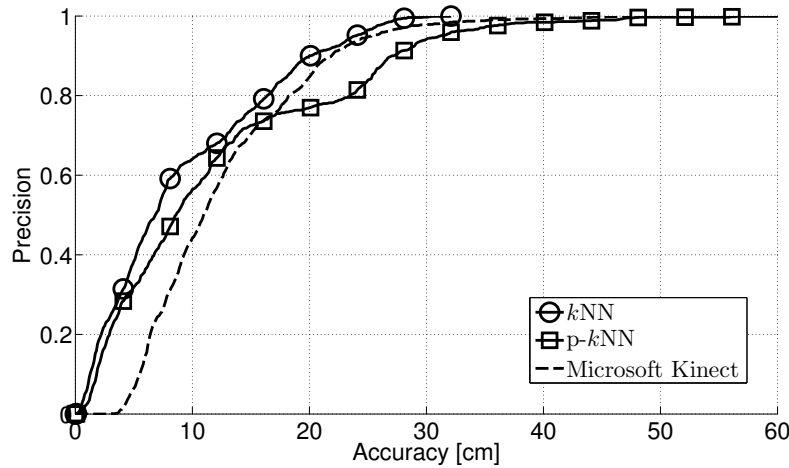
**Fig. 5** Precision, as a function of the accuracy, for both deterministic and probabilistic approaches. A comparison with the Kinect performance is also offered.

few transitional movements, which start and end at two of the trained fingerprint positions, were also executed (in order to test the system in *dynamic* conditions). For the purpose of comparison, a Microsoft Kinect system [2] was also used to perform arm motion capture.

In Fig. 5, the performance of the hybrid radio/accelerometric approach is shown, considering both deterministic and probabilistic approaches, along with that of Microsoft Kinect. It can be seen that the deterministic approach slightly outperforms the probabilistic one and the Kinect system, whereas the Kinect system outperforms the probabilistic approach for accuracy values higher than 15 cm (i.e., if a position error larger than 15 cm is accepted the Kinect system is more suitable than the probabilistic approach of our system).

### 2.3 The Inertial Approach

If one wants to avoid the use of fingerprinting and, thus, of radio signals, the only use of accelerometers is not sufficient to provide unambiguous information about the motion capture of the arm. Instead, a complete Inertial Measurement Unit (IMU) is generally used, which typically comprises a triaxial accelerometer and a triaxial gyroscope. Furthermore, a triaxial magnetometer can be included to measure the direction of the magnetic field in the surroundings of the sensor device (as a digital compass).

The inertial approach mainly relies on the use of an orientation estimation filter which must be able to continuously estimate and track the orientation of the sensor device, starting from the raw measurements of its acceleration (from the accelerom-

eter), angular velocity (from the gyroscope), and sensed magnetic field (from the magnetometer). The use of different sensors in the same device is mainly motivated by the fact that component sensors, if used independently, are prone to "intrinsic" (and unavoidable) problems. More precisely: accelerometers measure gravity acceleration, which is hardly discriminated from the linear acceleration of the device; gyroscopes can lead to a drift in the measurements due to implicit bias errors; magnetometers are very susceptible to magnetic disturbances of the local magnetic field (e.g., due to ferromagnetic materials in the proximity of the sensor).

The orientation of a device can be expressed, alternatively, in terms of: Euler angles (i.e., yaw, pitch, roll); rotation matrices; and quaternions [16]. Even though Euler angles give a more direct indication of how the device is rotated, quaternions are mostly used due to their robustness to unwanted phenomena (e.g., gimbal lock [16]) that can lead to singularities and ambiguities in the orientation estimation. In the last 20 years, many studies related to motion capture have led to the development of several types of orientation estimation filters. Furthermore, even if solutions based on Kalman filtering [14] have been mainly adopted in such studies [10, 19, 26, 28], in the last years many solutions have focused on complementary filtering techniques due to their simplicity and low computational requirements [20, 39, 40].

Provided that (i) each device is properly placed on each rigid body segment and (ii) its estimated orientation is correct, a kinematic chain thus appears. If a fixed origin is then chosen (e.g., the pelvis for full body motion capture, or a shoulder for arm motion capture), the positions of the extremes of all body segments can be easily derived from the estimated segments' orientations, leading to the reconstruction of the full posture of the user (or part of it in the case one wants to focus on a specific part). For ease of clarity, in Fig. 6 an illustrative representation of the reconstruction of the posture of the user, starting from the orientations of inertial devices attached to the user body segments, is shown only focusing on the upper body of the user.

The performance of the approach outlined above has been shown to be very accurate. Indeed, the estimated orientations of the body segments typically differ from the actual orientations of less than a degree in static conditions and of at most a few degrees in dynamic conditions, often leading to position errors in the order of a few centimeters.

## 3 A Comparative Overview

As previously discussed in Sect. 1, most of current state-of-the-art solutions for motion capture rely on optoelectronic or inertial approaches. However, the costs of an optoelectronic system and of a complete inertial suit are both very high. The only difference between these systems consists in the trade-off between accuracy and freedom of movement. Indeed, optoelectronic systems make errors in the order of a few millimeters, whereas errors with inertial systems are in the order of a few centimeters. Nevertheless, optoelectronic systems (and, more generally speaking, optical systems) are very sensitive to body occlusions and force the user to move in
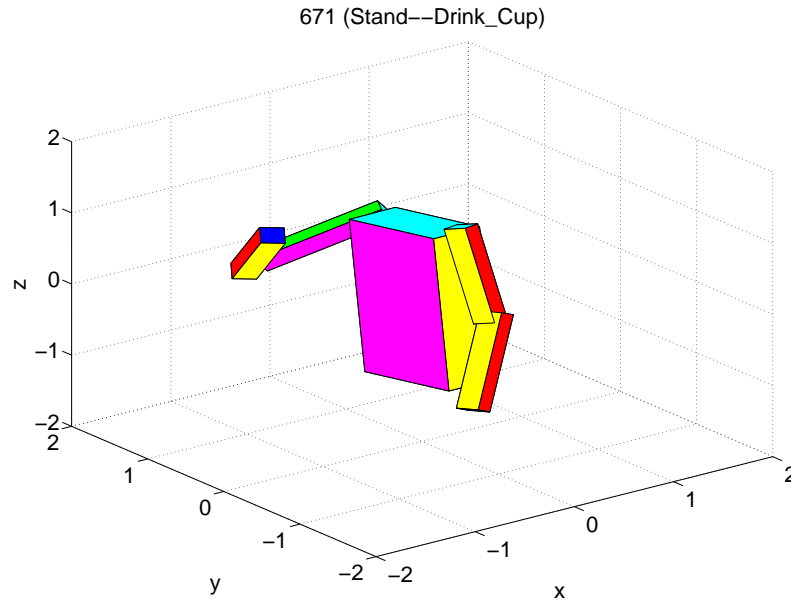
**Fig. 6** Illustrative representation of the reconstructed posture of the user upper body. As shown in the label, the user was drinking something from a cup.

a specific area. On the contrary, inertial systems allow users to move freely and do not suffer from occlusion problems.

Another major advantage of inertial systems is that they scale better in cost when the whole body motion capture is not necessary. Indeed, the cost of optoelectronic systems is mainly associated to high quality cameras whose use cannot be avoided, whereas inertial systems cost depends on the number of inertial devices that the user wants to use (and, thus, the number of body segments that the user wants to monitor). Furthermore, the cost of an inertial device mostly resides in the sensors cost. Therefore, the less sensors are required, the lower the cost is.

From the previous considerations, it is easy to understand that the transition between the three approaches described in Sect. 2 mainly owes to cost motivations. A device with only radio communication hardware (Sect. 2.1) costs less than the same device comprising also an accelerometer (Sect. 2.2), which, in turns, costs less than a device supplied with an accelerometer, a gyroscope, and a magnetometer (Sect. 2.3). However, one can also observe that the same transition allows to move toward systems with increasingly higher accuracy. Indeed, the radio-based approach is the less accurate, whereas the inertial approach is able to achieve the best performance. Similar considerations can be carried out with reference to the computational efforts and compactness of the various approaches. In particular, the radio-based approach requires limited computational efforts, whereas the radio/accelerometric hybrid approach and the inertial approach require—in this order—increasingly higher computational efforts. However, the inertial approach is self-contained and does

**Table 1** Comparative overview of the characteristics of the three presented approaches.

| | Component Technology | | |
|---|---|---|---|
| | **Radio** | **Radio/Accelerometric** | **Inertial** |
| **Accuracy** | Medium | Medium/High | High |
| **Cost** | Low | Medium | Medium/High |
| **Computational Efforts** | Low | Low/Medium | Medium |
| **Compactness (Self-containedness)** | Low | Low | High |

**Table 2** Overview of the advantages/disadvantages of the three presented approaches.

| | Component Technology | | |
|---|---|---|---|
| | **Radio** | **Radio/Accelerometric** | **Inertial** |
| **Advantages** | Very low cost | Low cost | Very accurate and independent of any external infrastructure |
| **Disadvantages** | External nodes are required (limited user motion freedom) and the estimate of the posture of the user is qualitative | External nodes are required (limited user motion freedom) | Higher cost and unable to provide the absolute position of the user |

not require any external infrastructure. Comparative overviews of the characteristics and of the advantages/disadvantages of the three approaches are summarized in Table 1 and Table 2, respectively.

Going deeper into the working principles of the three approaches, it can be of interest to highlight differences and similarities of the considered approaches. Concerning the radio approach (Sect. 2.1), it can be observed that the focus is on giving a qualitative (rather than quantitative) evaluation of the arm motion. Indeed, even if the interpolation between known trained postures allows to capture the arm motion in a continuous way, this approach can be perhaps more suitable to posture *classification*. The hybrid approach (Sect. 2.2), placing itself in the middle between the other two approaches, relies less on the interpolation. Indeed, through the use of a single sensor (i.e., the accelerometer), the hybrid approach makes use of a more reliable estimation of the device inclination and uses the radio signals component just to estimate the heading of the device. Finally, the inertial approach (Sect. 2.3)

completely abandons the contribution of the radio signals and estimate the position of the considered body segments starting only from the knowledge of the three-dimensional orientation of the device (estimated through a joint use of accelerometer, gyroscope, and magnetometer). This is also the reason why the first two approaches (Sect. 2.1 and Sect. 2.2) require the user not to move during the evaluation of the arm motion (or, at least, not to move the shoulder, being the origin of the reference system). Indeed, a relative difference between the position of the user and the positions of anchor nodes in his/her surroundings would invalidate the evaluation of the arm position. To this end, in [12] a hybrid placement of the anchor nodes also on the body user has been investigated, as an intermediate step toward the implementation of a fully portable system. Potentially, a future system design could exploit radio localization to directly estimate distances between (mobile) nodes placed only on the user body, possibly removing the distinction between anchor and target nodes (i.e., every node could be at the same time an anchor and a target node).

Concerning the use of the described approaches in specific applications, the performance results have proved that the inertial approach (Sect. 2.3) can be applied to a wide range of biomedical applications due to its high accuracy with respect to the others. The other approaches (Sect. 2.1 and Sect. 2.2) can instead be considered for health applications where it is relevant to discriminate discrete cases of arm motion (e.g., in rehabilitation after a surgery, if arm can be raised half-way or all the way).

Finally, a further development for the first two approaches (Sect. 2.1 and Sect. 2.2) can be that of exploiting the available environment infrastructure. For instance, if some fixed base stations (e.g., routers) are available in the room/building, they could be used as anchor nodes to capture the motion of the user in that room/building.

## 4 Case Studies

The previously presented approaches can be applied, for instance, in the following two case studies that will be straightforwardly referred to as *posture classification* and *motion capture*.

In *posture classification* applications, the aim is to determine the posture (i.e., the joints' positions) of a user body part (e.g., the arm), choosing between a predefined set of possible postures. In particular, the term *classification* specifically refers to a choice among a *discrete* set of postures. Therefore, it is not of interest to precisely estimate the positions of the joints, but a rough estimate is typically sufficient. This, however, typically implies that the system should be properly trained to recognize the user postures and discriminate between them. Finally, since the aim of posture classification is just to determine the posture of a user choosing among a discrete set of previously trained postures, posture classification could be considered in scenarios where a few cases need to be discriminated (e.g., while monitoring the recovery improvement after an arm surgery, it can be of interest to just evaluate whether the arm can be raised half-way or all the way).

On the other hand, in *motion capture* applications, the goal is to monitor and evaluate the precise movements of a user body part (e.g., the arm). This means recognizing and estimating the posture of the considered body part, without prior assumptions of possible postures that the body part can assume. In this case, the system does not need to be trained. However, since the problem is "unconstrained," it is not always easy to solve it and, in order to accurately estimate the body part posture, advanced signal processing techniques may be required.

At the end of this chapter, we will try to answer some common questions that can arise if one wants to apply the previously described techniques in the context of these two case studies.

## 5 Future directions

Due to the continuous advancements in the production of small and portable sensors, able to measure a wide variety of biophysiological signals, motion capture techniques and applications are probably experiencing, in these years, one of their most successful periods in terms of development and interest from a user perspective. The technology is almost mature and the first commercial applications are becoming month by month more known and used by the end users. The commercial cost, which now may still not be affordable for everyone, is continuously decreasing [1].

Especially in the gaming and biomedical fields, many applications based on motion capture and inertial sensors are already attracting the attention of gamers and doctors, respectively. In particular, in 5-10 years, the gaming scenario will probably be dominated by games where the user will interactively control virtual characters by using some easily wearable devices (at most 4-5 on the whole body, e.g., possible locations could include feet, ankles, wrists, and head [15, 17, 33]). The medical scenario will instead begin to adopt wearable sensors able to assess health of the patients by remotely measuring vital signs/parameters and allowing patients to freely stay at their homes still keeping a direct connection with their doctors [5, 7, 18, 23, 24, 25, 29, 37, 38, 41, 42].

Concerning the technologies described in this chapter, the living environments in the next years will be covered even more by radio devices able to communicate with sensors and devices placed on the user body, facilitating the application of the techniques previously described [35]. Also inertial and magnetic sensors will be further developed in order to become even more small, light, and less bulky. These advancements will then increase the "usability" gap between optical and inertial systems, making the latter more appealing for their flexibility and ease of use.

Finally, the technological developments foreseen above are also paving the way to the use of body sensor networking for more fundamental research activities, concerning the design of advanced biophysiological signal processing techniques able to detect, from posture and movement signals, the status of a person's pathological condition. In other words, the technological development of advanced posture

recognition techniques could allow to tackle relevant interdisciplinary (engineering/medical) research questions.

## 6 Conclusions

In this chapter, the importance and role of motion capture in mHealth applications has been investigated. To this end, three different motion capture approaches (i.e., radio-based, inertial, and radio/accelerometric) have been presented and described, in terms of performance, cost, and complexity. Furthermore, the advantages and limitations of all approaches have been discussed in a comparative way, characterizing their similarities and their differences and trying to highlight the most suitable application of each approach. Furthermore, some possible appealing extensions of the considered approaches have been also proposed.

## References

[1] Invensense. URL http://www.invensense.com/
[2] Kinect: markerless optical motion capture system from Microsoft. URL http://www.microsoft.com/en-us/kinectforwindows/
[3] Optoelectronic motion capture system from Vicon. URL http://www.vicon.com/
[4] Bahl P, Padmanabhan VN (2000) RADAR: an in-building RF-based user location and tracking system. In: Proc. IEEE Conf. on Computer Commun. (INFOCOM), Tel Aviv, Israel, vol 2, pp 775–784
[5] Bonato P (2010) Wearable sensors and systems. IEEE Eng Med Biol Mag 29(3):25–36
[6] Brodie M, Walmsley A, Page W (2008) Fusion motion capture: a prototype system using inertial measurement units and GPS for the biomechanical analysis of ski racing. Sports Technology 1(1):17–28
[7] Chen B, Patel S, Buckley T, Rednic R, McClure D, Shih L, Tarsy D, Welsh M, Bonato P (2011) A web-based system for home monitoring of patients with Parkinson's Disease using wearable sensors. IEEE Transactions on Biomedical Engineering 58(3):831–836
[8] Chen L, Wei H, Ferryman J (2013) A survey of human motion analysis using depth imagery. Pattern Recognition Letters 34(15):1995–2006
[9] Duda RO, Hart PE, Stork DG (2000) Pattern Classification and Scene Analysis. 2nd edition. Wiley-Interscience, New York, NY, USA
[10] Favre J, Jolles BM, Siegrist O, Aminian K (2006) Quaternion-based fusion of gyroscopes and accelerometers to improve 3D angle measurement. Electronics Letters 42(11):6125–614

[11] Giuberti M, Martalò M, Ferrari G (2011) Fingerprinting-based wireless 3D localization for motion capture applications. In: Proc. 1st ACM MobiHoc Workshop on Pervasive Wireless Healthcare (MobileHealth), Paris, France, pp 6.1–6.8

[12] Giuberti M, Martalò M, Ferrari G (2014) A hybrid radio/acclerometric approach to arm posture recognition. Journal of Ambient Intelligence and Smart Environments. Under revision. Available upon request.

[13] Kaemarungsi K (2005) Design of indoor positioning systems based on location fingerprinting technique. PhD thesis, University of Pittsburgh, Pittsburgh, URL http://etd.library.pitt.edu/ETD/available/etd-02232005-235903/unrestricted/dissertation28Feb05.pdf

[14] Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans of the ASME—Journal of Basic Eng 82(Series D):35–45

[15] Kim J, Seol Y, Lee J (2012) Realtime performance animation using sparse 3D motion sensors. In: International Conference on Motion in Games (MIG), pp 31–42

[16] Kuipers JB (1999) Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality. Princeton University Press, Princeton, NJ, USA

[17] Liu H, Wei X, Chai J, Ha I, Rhee T (2011) Realtime human motion control with a small number of inertial sensors. In: Proc. of Symposium on Interactive 3D Graphics and games (I3D), San Francisco, CA, USA, pp 133–140

[18] Lo G, Suresh AR, Stocco L, Gonzalez-Valenzuela S, Leung VCM (2011) A wireless sensor system for motion analysis of Parkinson's disease patients. In: 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, pp 372–375

[19] Luinge HJ, Veltink PH, Baten CTM (2007) Ambulatory measurement of arm orientation. Journal of Biomechanics 40(1):78–85

[20] Madgwick SOH (2010) An efficient orientation filter for inertial and inertial/magnetic sensor arrays. Tech. rep., Department of Mechanical Engineering, University of Bristol, Bristol, UK, URL http://www.x-io.co.uk/res/doc/madgwick_internal_report.pdf

[21] Martalò M, Giuberti M, Ferrari G (2011) Experimental investigation of wireless sensor networks for fingerprinting-based posture recognition. In: Riunione annuale 2011 del Gruppo nazionale Telecomunicazioni e Teoria dell'Informazione (GTTI), Messina and Taormina, Italy

[22] Moeslund TB, Hilton A, Krüger V (2006) A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 104(2):90–126

[23] Patel S, Park H, Bonato P, Chan L, Rodgers M (2012) A review of wearable sensors and systems with application in rehabilitation. Journal of NeuroEngineering and Rehabilitation 9(21):1–17

[24] Rahimi F, Duval C, Jog M, Bee C, South A, Jog M, Edwards R, Boissy P (2011) Capturing whole-body mobility of patients with Parkinson disease using inertial motion sensors: expected challenges and rewards. In: Proc. of the

33rd Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society (EMBS), Boston, MA, USA, pp 5833–5838

[25] Rodriguez-Silva DA, Gil-Castineira F, Gonzalez-Castano FJ, Duro RJ, Lopez-Pena F, Vales-Alonso J (2008) Human motion tracking and gait analysis: a brief review of current sensing systems and integration with intelligent environments. In: Proc. of IEEE Int. Conf. on Virtual Environments, Human-Computer Interfaces, and Measurement Systems (VECIMS), Instanbul, Turkey, pp 166–171

[26] Roetenberg D, Slycke PJ, Veltink PH (2007) Ambulatory position and orientation tracking fusing magnetic and inertial sensing. IEEE Trans on Biomedical Eng 54(5):883–890

[27] Roetenberg D, Luinge HJ, Slycke PJ (2009) Xsens MVN: Full 6DOF human motion tracking using miniature inertial sensors. Xsens Technologies B V

[28] Sabatini AM (2006) Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing. IEEE Trans on Biomedical Eng 53(7):1346–1356

[29] Salarian A (2006) Ambulatory monitoring of motor functions in patients with Parkinson's disease using kinematic sensors. PhD thesis, École polytechnique fédérale de Lausanne, Lausanne, Switzerland, URL http://biblion.epfl.ch/EPFL/theses/2006/3475/EPFL_TH3475.pdf

[30] Shepard D (1968) A two-dimensional interpolation function for irregularly spaced data. In: ACM National Conf.

[31] Swangmuang N, Krishnamurthy P (2008) Location fingerprint analyses toward efficient indoor positioning. In: IEEE Int. Conf. Pervasive Comp. and Commun. (PERCOM), Hong Kong, China, pp 100–109

[32] Tao Y, Hu H, Zhou H (2007) Integration of vision and inertial sensors for 3D arm motion tracking in home-based rehabilitation. The International Journal of Robotics Research 26(6):607–624

[33] Tautges J, Zinke A, Krüger B, Baumann J, Weber A, Helten T, Müller M, Seidel H, Eberhardt B (2011) Motion reconstruction using sparse accelerometer data. ACM Trans Graph 30(3):18:1–18:12

[34] US Wireless Corporation. RadioCamera Wireless Caller-Location System. URL http://www.directionsmag.com/companies/us-wireless-corp/93474

[35] Viani F, Robol F, Polo A, Rocca P, Oliveri G, Massa A (2013) Wireless architectures for heterogeneous sensing in smart home applications: Concepts and real implementation. Proceedings of the IEEE 101(11):2381–2396

[36] Vlasic D, Adelsberger R, Vannucci G, Barnwell J, Gross M, Matusik W, Popović J (2007) Practical motion capture in everyday surroundings. ACM Trans on Graphics 26(3):35.1–35.9

[37] Wong WY, Wong MS, Lo KH (2007) Clinical applications of sensors for human posture and movement analysis: a review. Prosthetics and Orthotics International 31(1):62–75

[38] Ying H, Schlösser M, Schnitzer A, Schäfer T, Schläfke ME, Leonhardt S, Schiek M (2011) Distributed intelligent sensor network for the rehabilitation of

Parkinson's patients. IEEE Trans on Information Technology in Biomedicine 15(2):268–276

[39] Young AD (2010) Wireless realtime motion tracking system using localised orientation estimation. PhD thesis, University of Edinburgh, Edinburgh, UK, URL http://www.xsens.com/images/stories/PDF/young_phd.pdf

[40] Yun X, Bachmann ER, McGhee RB (2008) A simplified quaternion-based algorithm for orientation estimation from earth gravity and magnetic field measurements. IEEE Trans on Instrumentation and Measurement 57(3):638–650

[41] Zheng H, Black ND, Harris ND (2005) Position-sensing technologies for movement analysis in stroke rehabilitation. Medical and Biological Engineering and Computing 43(4):413–420

[42] Zwartjes DGM, Heida T, van Vugt JPP, Geelen JAG, Veltink PH (2010) Ambulatory monitoring of activities and motor symptoms in Parkinson's Disease. IEEE Trans on Biomedical Eng 57(11):2778–2786

## Keywords

Motion capture, radio fingerprinting, inertial sensors, remote monitoring, body area network, body sensor network, sensor fusion.

## Questions and Answers

### *Case Study #1: "Posture Classification"*

As described in Sect. 4, in this scenario it is of interest to determine which posture is assuming a user body part (e.g., the arm), choosing between a predefined discrete set of possible postures.

- *Q1: Which approach would be preferable to use (among those presented in the chapter), in order to minimize the cost of the system but allowing a minimum level of acceptable performance?*
  *A1*: If the cost of the system should be minimized, the approach based on radio fingerprinting could be sufficient for the arm posture classification. Furthermore, the use of $k$NN and p-$k$NN with $k = 1$ should also be sufficient for such scenario. Indeed, it has been shown in the chapter that the estimation of the only fingerprint positions is very accurate also using the simplest version of radio fingerprinting. Inertial sensors are not required here.
- *Q2: Which kind of arm postures could be considered for the classification? How many devices would be required to be placed on the user body?*
  *A2*: The choice of the arm postures can be chosen by the doctor, recalling to
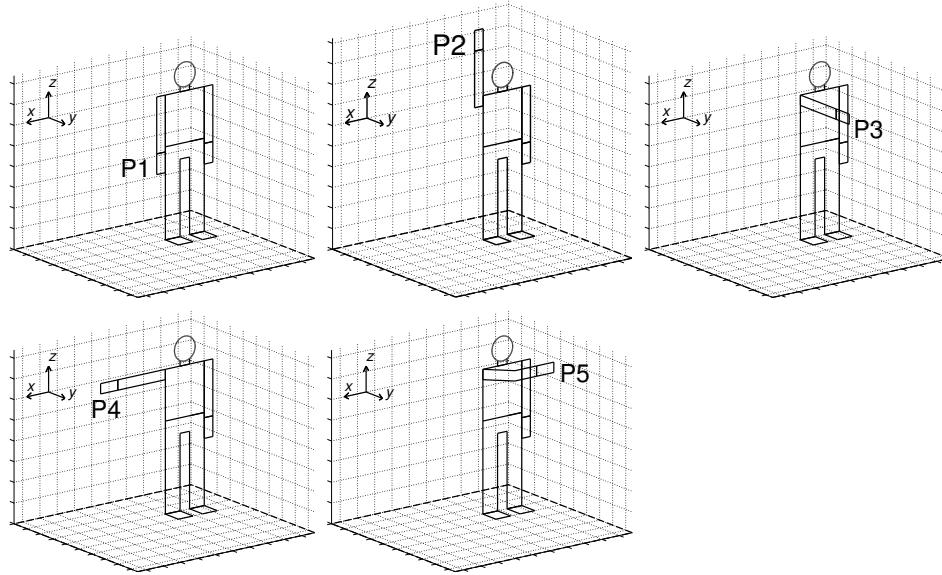
**Fig. 7** Representation of the 5 fingerprint positions (i.e., P1, P2, P3, P4, and P5) considered in the training phase.

properly train the system on the fingerprint positions associated to the chosen postures. Adjacent arm postures, however, should be chosen at a sufficiently large (spatial) distance in order to minimize the estimation error when classifying two different but close postures. For instance, in [11, 12, 21] the arm postures shown in Fig. 7 have been considered, leading to good classification performance. Finally, for these postures, even if the choice of two devices (i.e., on the upper arm and on the forearm) could be a more robust choice, one can consider to use only the node on the forearm, which is already associated with very distinct fingerprint positions (and, thus, arm postures). Observe that, if one is only interested in classifying different inclinations of the user arm (and its heading is not of interest), the best choice would be to use a device supplied with only an accelerometer.

- *Q3: In order to classify the posture of a leg, how should the system be changed?*
  *A3*: To this end, the devices should be placed on the user leg (e.g., one on the thigh and one on the shank) and during the session the user should pay attention not to move his/her hip (which will be now the origin of the reference system).

## *Case Study #2: "Motion Capture"*

As described in Sect. 4, in this scenario it is of interest to continuously monitor and evaluate the movements of a user body part (e.g., the arm).

- *Q1: Which approach would be preferable to use (among those presented in the chapter), in order to minimize the cost of the system but allowing a minimum level of acceptable performance?*
  *A1*: Even if almost every approach should be theoretically able to reconstruct the continuous movements of the arm (just discarding versions of $k$NN and p-$k$NN with $k = 1$), if one wants to obtain a minimum performance level, the approach based on the use of inertial and magnetic sensors would be the preferable. Indeed, in the chapter it has been mentioned that such approach could lead to errors in the order of only a few centimeters.
- *Q2: Which kind of arm movements could be evaluated using the previous approach? How many devices would be required to place on the user body?*
  *A2*: Every possible movement of the user arm could be evaluated. However, it must be clear that if one wants to estimate the motion of a specific arm segment a sensor device must be placed on that segment *and* on all the arm segments between it and the origin of the system (i.e., the shoulder). For instance, if one wants to estimate the motion of the hand, three sensor devices are necessary: one on the hand, one on the forearm, and one on the upper arm. Nevertheless, here the user is free to move in the room and no preliminary system training is needed.