Towards Cross-modal Comparison of Human Motion Data

Thomas Helten[1], Meinard Müller[1], Jochen Tautges[2],
Andreas Weber[2], and Hans-Peter Seidel[1]

[1]Max-Planck-Institut für Informatik
Campus E1.4, 66123 Saarbrücken, Germany
thelten@mpi-inf.mpg.de
[2]Bonn University, Institut für Informatik II
Friedrich-Ebert-Allee 144, 53113 Bonn, Germany

Abstract. Analyzing human motion data has become an important strand of research in many fields such as computer animation, sport sciences, and medicine. In this paper, we discuss various motion representations that originate from different sensor modalities and investigate their discriminative power in the context of motion identification and retrieval scenarios. As one main contribution, we introduce various mid-level motion representations that allow for comparing motion data in a cross-modal fashion. In particular, we show that certain low-dimensional feature representations derived from inertial sensors are suited for specifying high-dimensional motion data. Our evaluation shows that features based on directional information outperform purely acceleration based features in the context of motion retrieval scenarios.

1 Introduction

There are many ways for capturing and recording human motions including mechanical, magnetic, optical, and inertial devices. Each motion capturing (mocap) technology has its own strengths and weaknesses with regard to accuracy, expressiveness, and operating expenses, see [4,13] for an overview. For example, optical marker-based mocap systems typically provide high-quality motion data such as positional information given in joint coordinates or rotational information specified by joint angles. However, requiring an array of calibrated high-resolution cameras as well as special garment equipment, such systems are not only cost intensive but also impose limiting constraints on the actor and the recording environment. On the other side, in recent years low-cost inertial sensors, which can be easily attached to the body or even fit in a shoe, have become popular in computer game and sports applications [7,9]. Another use of inertial sensors is shown in [8], where the inertial sensor data is used to regularize marker-less tracking results. However, inertial information such as joint accelerations, angular velocities, or limb orientations, is often being of less expressive power and affected by noise.

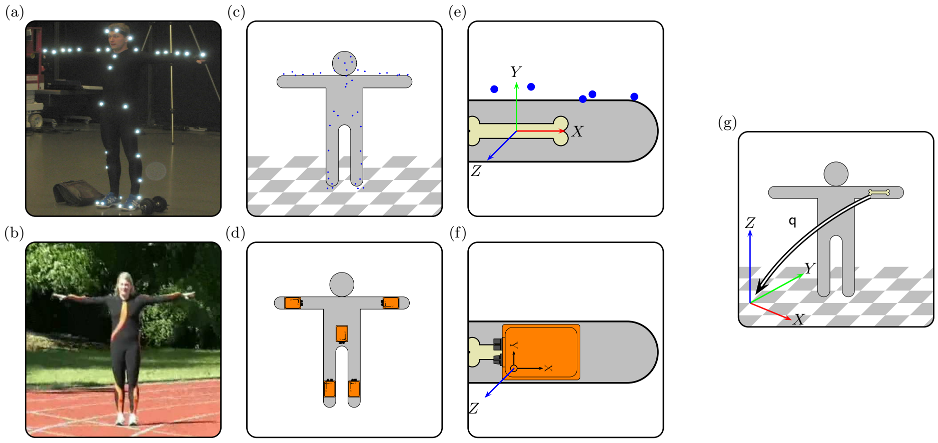


Fig. 1. (a) Actor wearing a suit with 41 retro-reflective markers as used by an optical mocap system. (b) Actress wearing a suit with 5 Xsens MTx Sensors. (c) Positions of 41 markers provided by the optical system. (d) Locations of the sensors. (e) Limbs' positions and orientations defined by the positions of markers. (f) Inertial sensors measuring the orientation of the limb they are attached to. (g) Limb orientation expressed with respect to a global coordinate system.

In this paper, we address the issue of cross-modal motion comparison investigating the expressiveness of various motion representations in the context of general motion identification and retrieval scenarios. As one main contribution, we introduce various mid-level feature representations that facilitate cross-modal comparison of various motion types. Here, the main challenge consists of finding a good trade-off between robustness and expressiveness: on the one hand, a mid-level representation has to be robustly deducible from the data outputted by different mocap systems; on the other hand, the representation has to contain enough information for discriminating motions within a certain application task. In particular, we show that certain low-dimensional orientation-based motion features are suited for accurately retrieving high-dimensional motion data as obtained from optical motion capturing.

The remainder of the paper is organized as follows. In Sect. 2, we describe different sensor modalities and discuss some of their properties. In particular, we go into more detail on acceleration and orientation data as obtained from recent inertial sensors. Then, in Sect. 3, we introduce various mid-level feature representations that can be derived from the different sensor modalities. In Sect. 4, we study the performance of these mid-level representations in the context of cross-modal motion retrieval. Finally, in Sect. 5 we conclude with an outlook on future work. Further related work is discussed in the respective sections.

2 Sensor Modalities

In this paper, we focus on two types of mocap systems, optical and inertial systems, which differ largely in acquisition cost, in the requirements on the recording conditions, and in the kind of data they provide. We now summarize some of the fundamental properties of such systems, while introducing several motion representations and fixing some notation.

2.1 Positional Motion Data

Optical marker-based mocap technology, as used in the passive marker-based Vicon MX system[1] or the active marker-based PhaseSpace system[2], allows for recording human motions with high precision. Here, the actor is equipped with a set of active or passive markers, which are tracked by an array of calibrated high-resolution cameras. From synchronously recorded 2D images of the marker positions, the system can then reconstruct 3D coordinates of marker positions or other skeletal kinematic chain representations. One particular strength of optical marker-based systems is that they provide positional motion data of high quality. However, requiring an array of calibrated high-resolution cameras as well as special garment equipment, such systems are cost intensive in acquisition and maintenance. Furthermore, many of the available optical mocap systems are vulnerable to bright lighting conditions thus posing additional constraints on the recording environment (e. g., illumination, volume, indoor). In our experiments, we use a set of 41 retro-reflective markers which are attached to an actor's suit at well defined locations following a fixed pattern, see Fig. 1 (a).

2.2 Inertial Motion Data

In contrast to marker-based reference systems, inertial sensors impose comparatively weak additional constraints on the overall recording setup with regard to location, recording volume, and illumination. Furthermore, inertial systems are relatively inexpensive as well as easy to operate and to maintain. Therefore, such sensors have become increasingly popular and are now widely used in many commercial products. On the downside, inertial sensors do not provide any high-qualitative positional data, but only accelerations and rate of turn data given in the sensor's local coordinate system. Note that these measured accelerations always contain, as one component, the acceleration caused by gravity. Therefore, the measured acceleration $\boldsymbol{a}$ can be thought of a superposition $\boldsymbol{a} = \overline{\mathsf{q}}[\boldsymbol{m} + \boldsymbol{g}]$ consisting of the gravity $\boldsymbol{g}$ and the actual acceleration $\boldsymbol{m}$ of the motion. Here, the quantity $\boldsymbol{a}$ is given in the sensors's local coordinate system, while $\boldsymbol{m}$ and $\boldsymbol{g}$ are given in the world coordinate system. The term $\overline{\mathsf{q}}[\cdot]$ represents the transformation from the global coordinate system to the sensor's local coordinate system (see below). This fact is often exploited in many portable devices such as recent

[1] `www.vicon.com`

[2] `www.phasespace.com`

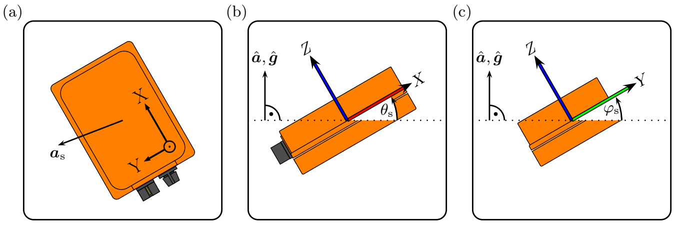

(a)
X
Y
$\boldsymbol{a}_\mathrm{s}$
(b)
$\hat{\boldsymbol{a}}, \hat{\boldsymbol{g}}$
Z
X
θ_s
(c)
$\hat{\boldsymbol{a}}, \hat{\boldsymbol{g}}$
Z
Y
φ_s

Fig. 2. Illustration of the different feature values. (a) Measured acceleration $\boldsymbol{a}_\mathrm{s}$ with respect to the sensors local coordinate system. (b) Pitch θ_s of a sensor with respect to the plane defined by $\hat{\boldsymbol{a}}$ respectively $\hat{\boldsymbol{g}}$. (c) Roll φ_s of a sensor with respect to the plane defined by $\hat{\boldsymbol{a}}$ respectively $\hat{\boldsymbol{g}}$.

mobile phones to calculate the device's orientation with respect to the canonical direction of gravity [2].

In the context of cross-modal comparison of optical and inertial data, one could integrate over the inertial data to obtain 3D positions. This, however, is not practical since inertial data is prone to noise leading to very poor positional data when being integrated [12]. Therefore, inertial data is often used indirectly to influence and control certain parameters within a motion generation engine. For example, inertial information may be used to identify and retrieve high-quality motions that were previously recorded by optical mocap systems [11]. Here, to make the 3D positional data comparable with inertial information, one obvious way is to suitably differentiate the 3D positional data to obtain velocities and accelerations. Such data, however, is very local in nature with respect to the temporal dimension thus making comparisons on this level susceptible to short-time artifacts and outliers. In the following sections, we investigate this issue in more detail and introduce mid-level representations that facilitate a more robust cross-modal comparison.

In our experiments, we use inertial sensors supplied by Xsens[3]. Each MTx unit contains an accelerometer, a rate gyro, as well as a magnet field sensor. These units combine the information of the contained sensors to calculate their full 3 degree of freedom (DOF) orientation q with respect to a global coordinate system, see e. g. [1,3]. In the following, we refer to such a combination of inertial and additional sensors as inertial unit. In order to express the orientation q we use rotations expressed as unit quaternions (see [10]). Each such quaternion defines a 3D rotation $\mathbb{R}^3 \to \mathbb{R}^3$, which we also refer to as q. Let $\mathsf{q}[\boldsymbol{x}]$ denote the rotated vector for a vector $\boldsymbol{x} \in \mathbb{R}^3$. The inverse rotation is referred to by $\overline{\mathsf{q}}$.

3 Feature Representations

In order to compare human motion data across different sensor modalities, one needs common mid-level representations that can be generated from the data

[3] `www.xsens.com`

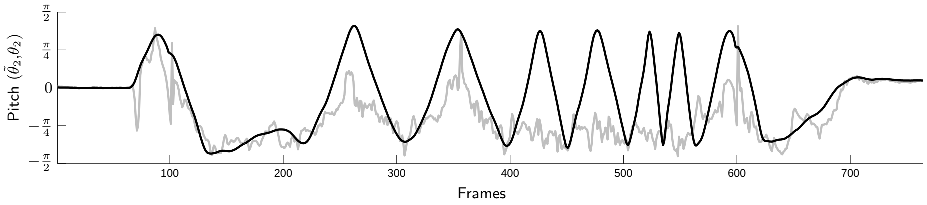


Fig. 3. Motion sequence consisting of six arm rotations, where the speed of the arm rotations increases with each repetition. The pitch of the left forearm is shown, calculated by using $\tilde{\theta}_2$ (gray) and θ_2 (black).

outputted by the different sensors. In the context of this paper, our goal is to retrieve full-body motions from a database containing motion data captured by an optical mocap system using 41 markers, where the query is given in form of a motion clip captured by five inertial sensors $s_1, \dots, s_5$ placed at the hip next to the spine (s_1) both lower arms (left s_2, right s_3) and both lower legs (left s_4, right s_5), see Fig. 1 (d). Since all information supplied by the five inertial sensors can be simulated using the 41 marker position (as shown in Sect. 3.1), we use features close to the inertial data as common mid-level representation.

3.1 Virtual Sensors

Local accelerations and directional information as provided by inertial sensors can also be defined from positional information coming from an optical mocap system. To this end, for a given inertial sensor fixed to a limb in a specific way, we use a suitable combination of markers to define the location and local coordinate system of a corresponding virtual sensor, see Fig. 1 (e). The orientation q of a virtual sensor is then the transformation from the local coordinate system to the global coordinate system (see Fig. 1 (g)), while the global acceleration $\boldsymbol{m}$ is obtained by double differentiation of the virtual sensor's global position. By adding the gravity $\boldsymbol{g}$ and transforming this quantity to the virtual sensor's local coordinate system using $\overline{\mathsf{q}}$ one finally gets the local acceleration $\boldsymbol{a} = \overline{\mathsf{q}}[\boldsymbol{m} + \boldsymbol{g}]$.

3.2 Local Acclerations

As a first simple feature representation, we directly use the local accelerations as outputted by the accelerometers. Using five inertial sensor units $s_1, \dots, s_5$, this results in five local accelerations $\boldsymbol{a}_{\mathrm{s}} \in \mathbb{R}^3$ for $s \in [1:5] := \{1, \dots, 5\}$. We then simply stack these five acceleration vectors to form a single vector

$$\boldsymbol{v}_{\boldsymbol{a}} = (\boldsymbol{a}_1^T, \dots, \boldsymbol{a}_5^T)^T \in \mathbb{R}^{15}. \tag{1}$$

3.3 Directions Relative to Acceleration

A more robust motion representation is obtained by measuring directions rather than magnitudes. To this end, we define a global up-direction using the direction of the gravity vector $\boldsymbol{g}$. We are now able to define a two degrees of freedom orientation of the sensor's local coordinate system relative to this global up-direction. Inspired by aviation, we call these two parameters *pitch* θ_s and *roll* φ_s, see Fig. 2. In many applications these quantities can be approximated using only the measured acceleration $\boldsymbol{a}_\mathrm{s}$. These approximations denoted by $\tilde{\theta}_\mathrm{s}$ and $\tilde{\varphi}_\mathrm{s}$, are defined as follows:

$$\hat{\boldsymbol{a}}_\mathrm{s} = \frac{\boldsymbol{a}_\mathrm{s}}{\|\boldsymbol{a}_\mathrm{s}\|}, \tag{2}$$

$$\tilde{\theta}_\mathrm{s} = \arccos \left\langle \hat{\boldsymbol{a}}_\mathrm{s}, (1,0,0)^T \right\rangle, \tag{3}$$

$$\tilde{\varphi}_\mathrm{s} = \arccos \left\langle \hat{\boldsymbol{a}}_\mathrm{s}, (0,1,0)^T \right\rangle. \tag{4}$$

Here, note that if the sensor's local Y-axis is perpendicular to the global up-direction, the pitch is determined by the rotation around the Y-axis. The resulting angle can be approximated by using an inner product between the X-axis and $\hat{\boldsymbol{a}}_\mathrm{s}$ approximating the up-direction, see Fig. 2 (b). Similarly, the roll can be derived from the inner product between the Y-axis and the upward direction, see Fig. 2 (c). We refer to the resulting pitch and roll features as *acceleration-based directional features*. Again, we stack these features for all five sensors $s_1, \ldots, s_5$ to form a single vector

$$\boldsymbol{v}_{\hat{\boldsymbol{a}}} = (\tilde{\theta}_1, \tilde{\varphi}_1, \ldots, \tilde{\theta}_5, \tilde{\varphi}_5)^T \in \mathbb{R}^{10}. \tag{5}$$

Here, pitch $\tilde{\theta}_\mathrm{s}$ and roll $\tilde{\varphi}_\mathrm{s}$ are calculated using $\boldsymbol{a}_\mathrm{s}$ as approximation for $\boldsymbol{g}$. Recall from Sect. 2.2 that each measured acceleration is a superposition $\boldsymbol{a}_\mathrm{s} = \overline{\mathsf{q}_\mathrm{s}}[\boldsymbol{m}_\mathrm{s}+\boldsymbol{g}]$. Thus $\tilde{\theta}_\mathrm{s}$ and $\tilde{\varphi}_\mathrm{s}$ are only good approximations if $\boldsymbol{m}_\mathrm{s}$ is negligible. However, for fast and dynamic motions, the component $\boldsymbol{m}_\mathrm{s}$ is large, which leads to corrupted pitch and roll values, see Fig. 3

3.4 Directions Relative to Gravity

To address the above mentioned problem, one needs to approximate the global upward direction in a more robust way—in particular during dynamic phases, where $\boldsymbol{m}_\mathrm{s}$ is not negligible. To achieve such an estimation, simple accelerometers do not suffice. We therefore use an inertial unit that outputs not only the local accelerations but also the sensor's orientation with respect to the global coordinate system, see Sect. 2.2. Then, the direction $\hat{\boldsymbol{g}}$ can be estimated by transforming the direction of the global Z-axis by means of the sensor's orientation q_s. More precisely, we define

$$\hat{\boldsymbol{g}}_\mathrm{s} = \overline{\mathsf{q}_\mathrm{s}} \left[(0,0,1)^T\right], \tag{6}$$

$$\theta_\mathrm{s} = \arccos \left\langle \hat{\boldsymbol{g}}_\mathrm{s}, (1,0,0)^T \right\rangle, \tag{7}$$

$$\varphi_\mathrm{s} = \arccos \left\langle \hat{\boldsymbol{g}}_\mathrm{s}, (0,1,0)^T \right\rangle. \tag{8}$$

As before, we stack the pitch and roll features for all five sensors $s_1, \ldots, s_5$ to form a single vector

$$\boldsymbol{v}_{\hat{\boldsymbol{g}}} = (\theta_1, \varphi_1, \ldots, \theta_5, \varphi_5)^T \in \mathbb{R}^{10}. \tag{9}$$

The components are referred to as *gravity-based directional features.* The values θ_{s} and φ_{s} exactly define (up to measurement errors) pitch and roll as introduced in Sect. 3.3. The improvements in the case of highly dynamic motions are illustrated by Fig. 3, which shows the values of $\tilde{\theta}_2$ and θ_2 over a motion sequence containing six arm rotations (between frames 210 and 575). Here, the arm rotations are performed at increasing speed, where the last rotation is performed almost three times faster than the first one. While θ_2 clearly shows the periodic fluctuation of the pitch during the rotation, $\tilde{\theta}_2$ fails to display any meaningful information when the motion becomes faster.

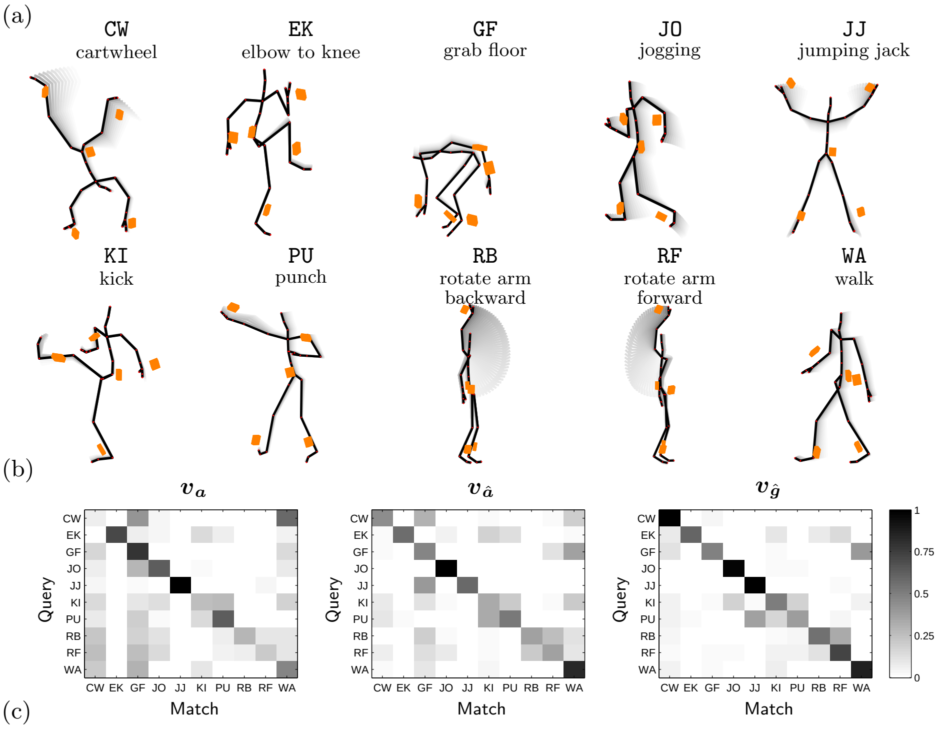
4 Cross-modal Comparison

In this section, we evaluate the feature representations in the context of a cross-modal retrieval scenario, where we search in a database which comprises high-dimensional 3D mocap while using low-dimensional inertial sensors as query input.To this end, we assembled two databases $\mathrm{DB}_{\mathrm{xse}}$ and $\mathrm{DB}_{\mathrm{c3d}}$. Each of the databases contains ten instances of the ten motion classes shown in Fig. 4 (a), which results in a total of 100 motion sequences per database. While the database $\mathrm{DB}_{\mathrm{xse}}$ was recorded using five inertial sensors set up as shown in Fig. 1 (d), the database $\mathrm{DB}_{\mathrm{c3d}}$ was assembled from excerpts of the HDM05 database which consists of high-quality motions recorded by a 12 camera Vicon optical mocap system, see [6]. Finally, we computed virtual sensors for $\mathrm{DB}_{\mathrm{c3d}}$ as described in Section 3.1 matching the sensor setup as used for $\mathrm{DB}_{\mathrm{xse}}$.

4.1 Class Confusion

Depending on the used feature representation, we now examine how well high-dimensional motion sequences in $\mathrm{DB}_{\mathrm{c3d}}$ can be characterized by low-dimensional sensor input from $\mathrm{DB}_{\mathrm{xse}}$. To this end, we rank the motion documents from $\mathrm{DB}_{\mathrm{c3d}}$ according to their similarity to a given query document from $\mathrm{DB}_{\mathrm{xse}}$. More precisely, we consider a document from $\mathrm{DB}_{\mathrm{c3d}}$ a match when it is an instance of the same motion class as the given query document from $\mathrm{DB}_{\mathrm{xse}}$. As similarity measure we use the classical dynamic time warping (DTW) distance described in [5], where, in our case, the highest ranked motion document has the smallest DTW distance. By considering the distribution of motion classes among the ten best-ranked documents one gets a good impression how the motion classes are confused under a given feature representation. A common means to visualize this are *confusion matrices*, which are shown for the three feature representations $\boldsymbol{v}_{\boldsymbol{a}}$, $\boldsymbol{v}_{\hat{\boldsymbol{a}}}$ and $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ in Fig. 4 (b). The rows of a confusion matrix represent the motion classes of the query, whereas the columns represent the motion classes of the

(a)



(c)

	CW	EK	GF	JO	JJ	KI	PU	RB	RF	WA	Ø
v_a	0.35	0.72	0.83	0.67	0.95	0.37	0.66	0.39	0.33	0.61	0.59
$v_{\hat{a}}$	0.57	0.64	0.62	1.00	0.68	0.42	0.62	0.43	0.49	0.86	0.63
$v_{\hat{g}}$	0.97	0.68	0.67	0.97	0.98	0.56	0.50	0.64	0.79	0.88	0.77

Fig. 4. (a) Motion classes used for the experiments in Sect. 4. **(b)** Confusion matrices (left) and true match distributions (right) of the three different feature representations. **(c)** Averaged maximal F-Measures for every feature representation and motion class. The last column shows for every feature representation the average over all motion classes.

ten best-ranked documents. Dark entries indicate a large percentage of a motion class, whereas light colors indicate a low percentage. For example, the matrices show that most of the motion classes are confused with the motion class CW (first column) when using the feature representation v_a. Here, the reason is that the motion class CW shows a lot of variance among the different motion instances even when performed by the same actor. In particular, the risk of confusion with the motion class CW is high for dynamic motions classes such as KI, PU, RB, and RF, because dynamic motions under the feature representation v_a have a very noisy character without much characteristic features. In contrast, using the directional feature representation $v_{\hat{g}}$ this confusion is reduced significantly.

4.2 F-measure

To further quantify the retrieval results, we use another measure from the retrieval domain referred to as *maximum F-measure*. Let k, $k \in [1:K]$ be the rank of a given document, where K is the maximum rank (in our case $K = 100$). Now, for every k, *precision* P_k and *recall* R_k are defined as $P_k := |T \cap M_k|/|M_k|$ and $R_k := |T \cap M_k|/|T|$. Here, M_k is the set of all documents up to rank k and T the set of all possible matches (in our case $|T| = 10$). Combining precision and recall values for a given rank k yields the (standard) F-measure $F_k := 2{\cdot}P_k{\cdot}R_k/(P_k{+}R_k)$. Now, the maximum F-measure is defined as $F := \max F_k, k \in [1:K]$. The table in Fig. 4 (c) shows the averaged maximum F-measure for each motion class,where the was calculated by averaging the maximum F-measures over all queries of each motion class, and every feature representation. Finally, the last column shows the average over all motion classes. The better a given feature representation discriminates a motion class against all other motion classes the larger is the corresponding entry in the table. It can be seen that the feature representation $\boldsymbol{v}_{\hat{\boldsymbol{a}}}$ is well suited to identify instances of motion class `JO` (1.00), whereas the feature representation $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ performs particularly well for the motion classes `CW` (0.97), `JO` (0.97), and `JJ` (0.98). Furthermore, the identification rates for the class `CW` show a drastic improvement under the feature representation $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ (0.97) in comparison to $\boldsymbol{v}_{\boldsymbol{a}}$ (0.35). Also, the arm rotations `RB` and `RF` are much better characterized under the feature representation $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ (0.64 and 0.79) compared to the acceleration based feature representations $\boldsymbol{v}_{\boldsymbol{a}}$ (0.39 and 0.33) and $\boldsymbol{v}_{\hat{\boldsymbol{a}}}$ (0.43 and 0.492). Interestingly, there are some exceptions where $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ does not outperform the other two feature representations. For example, in case of motion class `PU`, $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ (0.50) is worse compared to $\boldsymbol{v}_{\hat{\boldsymbol{a}}}$ (0.62) and $\boldsymbol{v}_{\boldsymbol{a}}$ (0.66). Here, on the one hand, the orientations of both arms—including roll and pitch—shows large variations among the actors. On the other hand, all punching motion exhibit characteristic peaks in the acceleration data, which can be captured particulary well by $\boldsymbol{v}_{\boldsymbol{a}}$. However, in general, one can notice that $\boldsymbol{v}_{\hat{\boldsymbol{g}}}$ is much better suited to identify most motion classes than the feature representations $\boldsymbol{v}_{\boldsymbol{a}}$ and $\boldsymbol{v}_{\hat{\boldsymbol{a}}}$.

5 Conclusions

In this paper, we have presented a systematic analysis of various feature representations in the context of a cross-modal retrieval scenario, where inertial-based query motions are used to retrieve high-quality optical mocap data. Because of the increasing relevance of motion sensors for monitoring and entertainment purposes, the fusion of various sensor modalities as well as cross-domain motion analysis and synthesis will further gain in importance. For example, first approaches have been presented that allow for identifying high-quality 3D human motions from sparse inertial sensor input [11]. The reconstruction of high-quality 3D human motions using database knowledge has become a major principle used in computer animation and the gaming industry. Here, our analysis results and methods constitute a suitable foundation for estimating the performance of the various motion representations. As one main result, we showed that directional features relating the sensor to the direction of gravity outperform purely acceleration-based features. In particular, it turns out that rate-of-turn data is necessary to enhance the roll and pitch estimates in the case of dynamic, fast changing motions. In this context, we plan to investigate which kind of sensor input in combination with pre-recorded motion data is necessary for reconstructing full body human motions.

References

1. Kemp, B., Janssen, A.J.M.W., van der Kamp, B.: Body position can be monitored in 3d using miniature accelerometers and earth-magnetic field sensors. Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control 109(6), 484–488 (1998)
2. Lee, J., Ha, I.: Real-time motion capture for a human body using accelerometers. Robotica 19(06), 601–610 (2001)
3. Luinge, H.J., Veltink, P.H.: Measuring orientation of human body segments using miniature gyroscopes and accelerometers. Medical and Biological Engineering and Computing 43(2), 273–282 (2005), `http://doc.utwente.nl/61405/`
4. Maiocchi, R.: 3-d character animation using motion capture pp. 10–39 (1996)
5. Müller, M., Röder, T.: Motion templates for automatic classification and retrieval of motion capture data. In: Proc. ACM SCA. pp. 137–146. ACM Press (2006)
6. Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., Weber, A.: Documentation: Mocap Database HDM05. Computer Graphics Technical Report CG-2007-2, Universität Bonn (Jun 2007), `http://www.mpi-inf.mpg.de/resources/HDM05`
7. Ohgi, Y., Ichikawa, H., Miyaji, C.: Microcomputer-based acceleration sensor device for swimming stroke monitoring. JSME International Journal Series C Mechanical Systems, Machine Elements and Manufacturing 45(4), 960–966 (2002)
8. Pons-Moll, G., Baak, A., Helten, T., Müller, M., Seidel, H.P., Rosenhahn, B.: Multisensor-fusion for 3d full-body human motion capture. In: Proc. IEEE CVPR, to appear. vol. 0, pp. 663–670 (Jun 2010)
9. Sabatini, A., Martelloni, C., Scapellato, S., Cavallo, F.: Assessment of walking features from foot inertial sensing. IEEE Transactions on Biomedical Engineering 52(3), 486–494 (2005)
10. Shoemake, K.: Animating rotation with quaternion curves. ACM SIGGRAPH Computer Graphics 19(3), 245–254 (Jul 1985)
11. Slyper, R., Hodgins, J.: Action capture with accelerometers. In: Proc. ACM SCA (Jul 2008)
12. Thong, Y.K., Woolfson, M.S., Crowe, J.A., Hayes-Gill, B.R., Jones, D.A.: Numerical double integration of acceleration measurements in noise. Measurement 36(1), 73–92 (2004)
13. Zheng, H., Black, N., Harris, N.: Position-sensing technologies for movement analysis in stroke rehabilitation. Medical and Biological Engineering and Computing 43, 413–420 (2005)