An Action Recognition Method Based on Wearable Sensors

Fuliang Ma[1], Jing Tan[1], Xiubing Liu[1], Huiqiang Wang[1], Guangsheng Feng[1](✉), Bingyang Li[1], Hongwu Lv[1], Junyu Lin[2], and Mao Tang[3]

[1] College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China
fengguangsheng@hrbeu.edu.cn
[2] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[3] Science and Technology Resource Sharing Service Center of Heilongjiang, Harbin 150001, China

Abstract. In the field of human action recognition, some existing works are mainly focused on macro actions, e.g., the requirements for action recognition is walking or jumping, while others are concentrated on micro actions, e.g., hand waving or leg raising. However, existing works rarely consider the recognition effect of different sensor wearing schemes with various requirements. In this work, the influences of the wearing scheme on action recognition effect are taken into account, a universal action recognition method to adapt different recognition requirements is developed. First, we present an action layered verification model which includes static action layer, dynamic action layer and joint presentation layer, which is used to provide an optional wearing scheme for each layer and to prevent wrong classification problems. Second, we verify the recognition effect of various wearing schemes under different layers. Finally, an action recognition method based on decision tree is introduced to adapt different requirements. The experiments show that the proposed method achieves a desirable recognition effect in comparison to existing ones.

Keywords: Action recognition · Wearable sensors · Wearing scheme

1 Introduction

Human action recognition is a hot topic in the field of human-computer interaction (HCI) and has received widespread attentions as the techniques of HCI and communication are making continuous improvement [1]. Some existing works are mainly focused on macro action recognition, e.g., the requirements for action recognition is walking or jumping, while others are on micro actions, e.g., hand waving or leg raising [2–5]. Moreover, the earlier technologies to recognize actions by image analysis, captured by the pre-installed cameras [6–9], which is severely affected by the camera accuracy and also incurs privacy concerns. Recently, the wearable sensor-based recognition are coming into interest due to the merits of the sensors, including small size, easy wearing, privacy-protecting, etc. [10–12].

J. Zheng et al. (Eds.): ADHOCNETS 2018, LNICST 258, pp. 202–211, 2019.
https://doi.org/10.1007/978-3-030-05888-3_19

To recognize different kinds of human actions, placing a variety of sensors on different positions is usually employed, e.g., accelerometers on the abdomen for elderly fall detection [13], acceleration and pressure sensors on the sole for recognizing walking, sitting and standing [2], accelerometers in thigh pockets for recognizing athlete's swimming [14], mobile phones with acceleration sensors on leg for recognizing climbing stairs [3], etc. The influences of sensor positions on action recognition are also considered by some researchers. For example, the accelerometer placed on chest has slight advantages on gesture recognition and fall detection [15]. Moreover, the work [16] studies the sensor position calculation for improving the recognition accuracy. In the above studies, the positions and numbers of sensors used to recognize actions are different, which shows that the sensor arrangement is significantly crucial. Although these studies have achieved good results in specific areas, however, the recognition effect of different sensor wearing schemes under different requirements is rarely considered.

In reality, people have different requirements for the action recognition, such as static, dynamic and mixed actions, which is the main problem considered in this work, i.e., designing an action recognition method to adapt different recognition requirements. First, a layered verification model is developed to distinguish different layer actions and verify the recognition effect of the sensor wearing schemes under different layers. Then, we present a universal wearing scheme for actions by comparing the wearing schemes at different layers. Finally, an action recognition method is designed to adapt different requirements.

2 System Model

Figure 1 shows the proposed recognition framework, including data processing, action layered verification model and action recognition model. In the framework, a layered verification model is designed based on the random forest classifier, which can be used to examine various wearing schemes of action recognition on different layers. On this basis, a universal action recognition method after analyzing the sensor for wearing scheme on different layers is developed.

2.1 Data Processing

2.1.1 Data Collection

The Ubisense positioning platform and positioning tags are used for data collection [17]. The platform consists of three components: sensors, positioning tags and positioning platform iLocateTRM, where the positioning tags transmit position information to the sensors via Ultra Wideband (UWB) pulse signal. After receiving the signal, the sensor adopts TDOA and AOA positioning. The algorithm analyzes the tags location and finally transmits it to the iLocate server via wired Ethernet. In an indoor environment, the platform can stably achieve a 3D positioning accuracy of 15 cm.

We place position sensors at 10 positions, including chest (P1), abdomen (P2), left upper arm (P3), left forearm (P4), right upper arm (P5), right forearm (P6), left thigh (P7), left lower leg (P8), right thigh (P9) and right lower leg (P10). Then we build a data set which includes crouch(A1), lying (A2), sitting (A3), standing (A4), walking (A5), tiptoe (stepping on the tip of the toe, (A6), body turn (A7), squat downward (A8), bending arm (A9), raising hand (A10), lifting the leg (A11) and lifting heavy objects (A12). The data set is collected from ten males and ten females. The participants range in height from 1.60 m to 1.78 m, and their weights vary from 50 kg to 85 kg. We continuously collect each action for ten minutes. The Ubisense positioning platform reports the position information of each tags at the frequency of 10 Hz, and the data is saved in the format of <ID, T, X, Y, Z> .

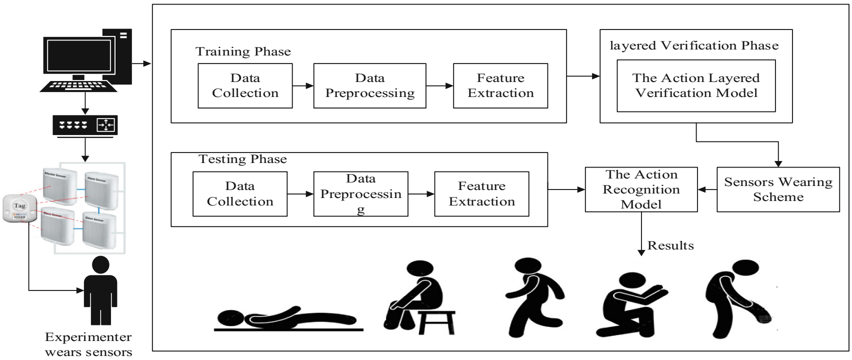

Fig. 1. Action recognition framework.

2.1.2 Data Preprocessing and Feature Extraction

Considering the unavailable noise of raw data, a median filter with a window size of 3 is used. The overlapping time window is a common way to extract features from a time-based data stream. In this paper, we verify the window of 1 to 2 s in consideration of the completeness of actions. Finally the time window size is determined as 1.4 s.

We consider three types of features that include action features, relative features, and statistical features, as shown in Table 1, in which the relative features represent the relationship between any two sensors.

Table 1. Features in detail.

	Features
Motion features	Speed, acceleration, displacement, displacement in the time window, height
Relative features	Relative speed, relative acceleration, relative displacement, relative height
Statistical features	Mean of displacement, standard deviation of displacement, mean of displacement between two sensors, standard deviation of displacement between two sensors

2.2 The Action Layered Verification Model

We design static action layer, dynamic action layer, and joint presentation layer according to the difference in action amplitude and the relationship between limbs. The layers are shown in Table 2.

Table 2. The layers.

ID	Layer	Action	ID	Layer	Action
A1	Static action layer	Crouch	A7	Dynamic action layer	Body turn
A2		Lying	A8		Squat downward
A3		Sitting	A9	Joint presentation layer	Bending arm
A4		Standing	A10		Raising hand
A5	Dynamic action layer	Walking	A11		Lifting the leg
A6		Tiptoe	A12		Lifting heavy objects

The action layered verification model serves the action recognition model. In previous studies, sensor wearing schemes are subjective and empirical. The action layered verification model attempts to explain the effect of the position more objectively. The model provides an optional wearing solution for each layer of action. Then we can adjust the wearing schemes to adapt the requirements (i.e., the requirement for action recognition in the traditional model is to recognize only static actions such as sitting, standing and lying.) of different action recognition.

The data used in this paper is processed during the training phase. First, we use the random forest classifier to layer it in accordance with Table 2. Then the separate classifier is designed for each layer. Under each classifier, we extract a combination of different wearing schemes from the data set to verify a better wearing scheme for each layer. We use four or less sensors to carry out this experiment here, because too many sensors may cause signal interference and waste of resources.

2.3 The Action Recognition Method

The action recognition method includes an action layered model and an action classification model. In the initial stage, the action layered model is used to avoid wrong action classification. Then action classification models are designed for each layer. In action recognition method, the training data set is D = {X1, X2, X3, …, Xn}, and feature set is A = {A1, A2, A3, .., Av}. D belongs to a group of classes C = {C1, C2, C3, …, Cw}, and also belongs to a group of classes L = {L1, L2, …, Ls}. Here, C is the class of the action, and L is the layer of the action.

The random forest is used as the action layered model, which includes model training and action layering. The processes of model training is as follows: (1) select N samples from the training data set by putting back random samples. (2) Use K features selected randomly to establish a decision tree. (3) Repeat the above two steps m times to generate m decision trees to form a random forest. The processes of action layering is: (1) for the test data, all decision trees are classified one by one. Then we vote to determine the layering result. (2) The layering result L is added as a feature to the raw training data set to form a new training set for action classification. The feature set is represented as A = {A_1, A_2, A_3, .., A_v, L}.

In action classification model, L is only used to distinguish the layer to which the action belongs. A C4.5 decision tree based classification model is created for each action layer L. The action classification model is constructed as follows: Select the best feature and segmentation point according to the information gain rate, and then split the root node into several sub-nodes according to the best feature and segmentation point. Second, split the sub-node into several sub-segments according to the best feature and segmentation point similarly. The child nodes are recursively split until the recursion end condition is satisfied. There are several cases of recursion end condition: (1) the sample categories in the child nodes are of the same category, (2) the attribute is an empty set, and (3) the feature information gain rate is less than the threshold.

3 Experiment Analysis

In this part, we first analyze and verify each layer of wearing scheme. Then we combine the characteristics of each layer in wearable scheme to verify the action recognition method that applies to all actions. Finally, we validate our scheme on Naive Bayes (NB), Support Vector Machine (SVM) and Artificial Neural network (ANN).

3.1 Action Layered

Table 3 shows that the effects of layered performs very well (99%) and explains the feasibility of the model we proposed. We use the following metrics to show the classification effect: Precision (P), Recall ratio (Recall/R), False Positive Rate (FP Rate), F-Measure. F-Measure is defined as follow.

$$F - Measure = 2 * \frac{P * R}{P + R} \tag{1}$$

Table 3. Layered effects of random forest classifier.

Action layer	Precision	Recall	FP rate	F-measure
Static action layer	0.997	0.996	0.001	0.996
Dynamic action layer	0.990	0.986	0.005	0.988
Joint presentation layer	0.983	0.988	0.009	0.985
Average	0.990	0.990	0.006	0.990

3.2 Layered Verification

3.2.1 Static Action Layer

Static action layer contains four actions A1-A4. We train classifiers for each wearing scheme separately. Figure 2 outlines the results that the positions of wearable sensor have great effect on the recognition quality of the static action layers. It shows that chest and abdomen have the highest recognition rate (96%), arm (P3-P6) performs well (91%) and leg (P7-P10) is the worst (74.8%/67.8%). We examine the feature set and find that the height feature provides useful information, and it performs well when wearable sensors locate in upper body.

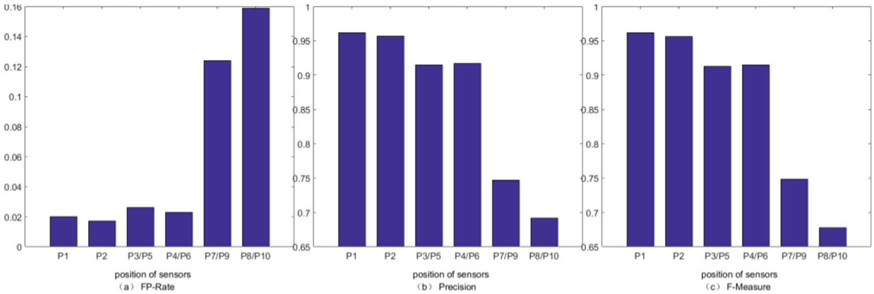

Fig. 2. Static action layers: comparison of performance under different wearing scheme.

3.2.2 Dynamic Action Layers

Dynamic action layers contain four actions A5-A8. We train classifiers for each wearing scheme separately. Figure 3 shows the performance comparison using single sensor under dynamic action layer. The result ($\leq$ 70%) is not satisfactory when we use a single sensor to recognize the action of this layer. Therefore, we combine sensor in the best position P1 with other positions, which has a satisfactory effect (Fig. 4). The chest, arm and thigh have a higher recognition rate ($\geq$ 88.6%), and the standard combination of chest and crus is relatively poor (85.9%). The results show that the increase in the number of sensors and the introduction of relative relationship can improve the recognition performance effectively.

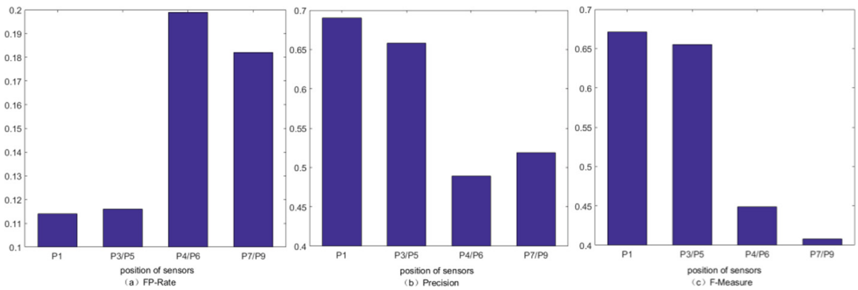

Fig. 3. Dynamic action layer: performance comparison using a single sensor.

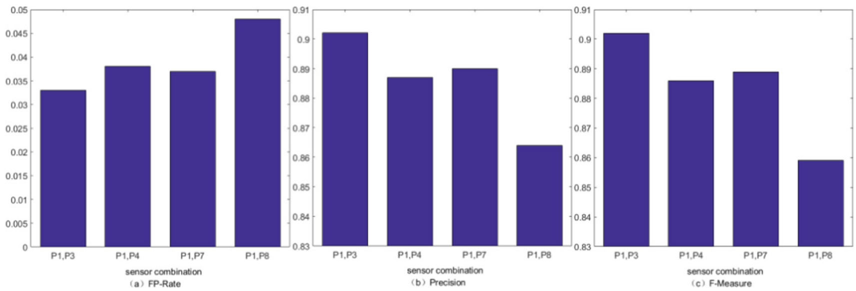

Fig. 4. Dynamic action layer: performance comparison using a combination of two sensors.

3.2.3 Joint Presentation Layers

Joint presentation layer contains four actions A9-A12. We train classifiers for each wearing scheme separately. Using a single sensor cannot represent the complete action of the joint presentation layer, such as raise hand. We directly use the combination of torso, upper extremity and lower extremity position to recognize the action of the joint presentation layer. Because the action includes both upper and lower extremities. Figure 5 shows that combination of three sensors represents the joint presentation layer action well ($\geq$ 80%). Besides, the combination of the left arm (P3, P4) and lower limbs (best 91.7%) is almost higher than that of right arm (P5, P6) and lower limbs. To ensure stability, it is recommended to use the sensor combination on the left body when recognizing the action of the joint presentation layer.

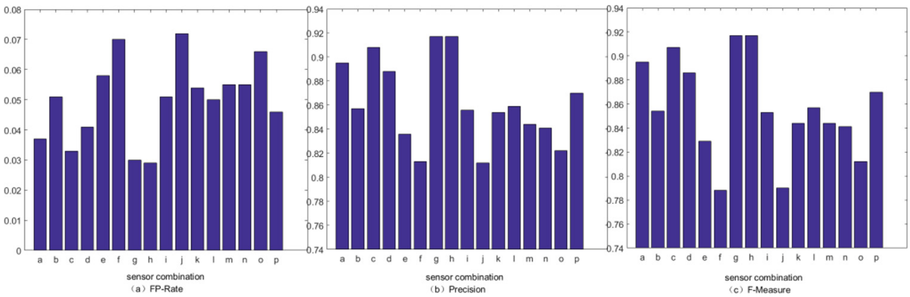

Fig. 5. Joint presentation layers: performance comparison using a combination of three sensors. (Remark: a:{P1, P3, P7}, b:{P1, P3, P8}, c:{P1, P4, P7}, d:{P1, P4, P8}, e:{P1, P3, P9}, f:{P1, P3, P10}, g:{P1.P4, P9}, h:{P1, P4, P10}, i:{P1, P5, P7}, j:{P1, P5, P8}, k:{P1, P5, P9}, l:{P1, P5, P10}, m:{P1, P6, P7}, n:{P1, P6, P8}, o:{P1, P6, P9}, p:{P1, P6, P10}).

3.3 Universal Recognition Method

In the above experiments, we validate the sensor wearing scheme for each layer. We find that there are similarities among each layer's wearing scheme. P1 is an absolutely necessary position in all well-performed wearing schemes. P3, P4, and P7 also perform well in the dynamic action layer and the joint presentation layer. Besides, we find that the combination of {P1, P3, P7} is the best when we choose three sensors in {P1, P3, P4, P7}, and P1 must be chosen. The result is shown in Table 4.

Table 4. The confusion matrix under the specific wearing scheme.

	A1	A2	A3	A4	A5	A6	A7	A8	9A	A10	A11	A12
A1	561	3	4	0	2	0	0	8	0	0	5	0
A2	20	558	0	0	0	0	0	0	0	0	0	0
A3	2	0	550	0	0	0	0	18	0	0	13	0
A4	0	0	0	506	4	27	18	2	27	1	0	10
A5	2	0	1	0	523	10	27	10	0	5	0	5
A6	0	0	0	29	10	455	9	5	27	5	0	25
A7	0	0	0	15	19	16	460	4	21	44	0	12
A8	5	1	12	2	7	3	3	542	0	8	8	6
A9	0	0	0	9	0	23	20	0	472	9	0	40
A10	0	0	0	6	6	5	38	5	5	513	0	15
A11	1	0	15	0	0	0	0	15	0	0	553	0
A12	2	0	0	15	7	31	13	2	102	14	0	399
Precision	0.95	0.99	0.95	0.87	0.91	0.80	0.78	0.89	0.72	0.86	0.96	0.78

The table shows that the static action layer except standing (A4) has well performance under our wearing scheme ($\geq$ 95%). Standing is wrongly classified into {A6, A7, A9, A12}, and these actions are not recognized well ($\leq$ 80%). The reason is that these actions are all based on standing with tiny differences. Besides, we demonstrate that increasing sensors can improve the recognition performance effectively. In Table 5, we analyze the wearing schemes with 4 sensors and give the confusion matrix of the best wearing scheme {P1, P3, P5, P7}.

Table 5. Confusion matrix of the best wearing scheme.

	A1	A2	A3	A4	A5	A6	A7	A8	9A	A10	A11	A12
A1	568	1	1	0	1	0	0	10	0	0	1	0
A2	16	575	0	0	0	0	0	0	0	0	0	0
A3	1	0	574	0	0	0	0	7	0	0	11	0
A4	0	0	0	534	4	18	11	2	2	1	0	8
A5	2	0	1	6	518	11	26	12	0	1	0	5
A6	0	0	0	13	9	513	11	7	22	4	0	26
A7	0	0	0	15	20	7	466	5	4	19	0	23
A8	4	1	7	2	7	6	4	537	1	5	6	5
A9	0	0	0	2	0	8	9	0	522	9	0	52
A10	0	0	0	4	1	9	36	6	6	540	0	9
A11	3	0	5	0	0	0	0	9	0	0	568	0
A12	0	0	0	7	3	26	15	5	62	12	0	474
Precision	0.96	0.99	0.98	0.92	0.92	0.86	0.81	0.90	0.84	0.91	0.97	0.79

3.4 Comparison with Other Classification Methods

In order to show the benefits of our wearing scheme, we compare the performance under the classifiers including DT, ANN, NB and SVM. Figure 6 shows that our wearing scheme outperforms other schemes. Meanwhile in our wearing scheme, the other classifiers perform worse than our DT classifier. Figure 6 shows clearly that DT (90.3%) outperforms other classifiers.

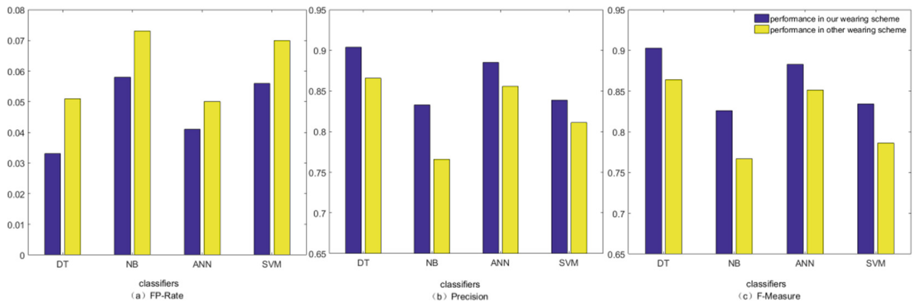

Fig. 6. Comparisons of proposed wearing scheme and classifier.

4 Conclusion and Future Work

In order to adapt different requirements for action recognition, we propose an action layered verification model. It is based on the random forest classifier, which can achieve 99% accuracy of layering. Then we verify wearing schemes of each layer. The experiments show that sensors in chest position provide reliable information in the static action layer, and the recognition rate reaches 96%. It fails to achieve a satisfactory result when we use single sensor to recognize the actions of dynamic action layer. However, using a combination of two sensors located on the chest and arm achieves a recognition rate more than 89%. Besides, three sensors are required at least to recognize the action of the joint presentation layer. Because using a combination of two sensors on the left side of the body and chest, the recognition rate only reaches 86%.

Subsequently, experiments are conducted to identify the overall actions. After a comprehensive analysis of the wearing scheme each layer, we propose an action recognition model based on four sensors. The recognition rate reaches 90.3%. Meanwhile, we verify our wearing scheme on other classifiers. The experiments show that our wearing scheme is also applicable to other classifiers, and the performance of action recognition improves by our wearing scheme.

In the future work, we will find more effective approaches to recognize similar actions and build a complete action recognition system.

Acknowledgement. This work is supported by the Natural Science Foundation of China (No. 615 02118), the Natural Science Foundation of Heilongjiang Province in China (No. F2016009), the Fundamental Research Fund for the Central Universities in China (No. HEUCF180602 and HEUCFM180604) and the National Science and Technology Major Project (No. 2016ZX0 3001023-005).

References

1. Moayedi, F., Azimifar, Z.S., Boostani, R.: Structured sparse representation for human action recognition. Neurocomputing **161**(C), 38–46 (2015)
2. Sazonov, E.S., Fulk, G., Hill, J., Browning, R.: Monitoring of posture allocations and activities by a shoe-based wearable sensor. IEEE Trans. Biomed. Eng. **58**(4), 983–990 (2011)
3. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. Acm Sigkdd Explor. Newsl. **12**(2), 74–82 (2016)
4. Liu, Y., Nie, L., Liu, L., Rosenblum, D.S.: From action to activity: sensor-based activity recognition. Neurocomputing **181**, 108–115 (2016)
5. Zhang, M., Sawchuk, A.A.: Human daily activity recognition with sparse representation using wearable sensors. IEEE J Biomed Health Inform **17**(3), 553–560 (2013)
6. Ullah, A., Ahmad, J., Muhammad, K.: Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access **PP**(99), 1 (2017)
7. Panagiotakis, C., Papoutsakis, K., Argyros, A.: A graph-based approach for detecting common actions in motion capture data and videos. Pattern Recogn. **79**, 1–11 (2018)
8. Alfaro, A., Mery, D., Soto, A.: Action recognition in video using sparse coding and relative features, pp. 2688–2697 (2016)
9. Rautaray, S.S., Agrawal, A.: Vision based hand gesture recognition for human computer interaction: a survey. Artif. Intell. Rev. **43**(1), 1–54 (2015)
10. Saner, H.: Wearable sensors for assisted living in elderly people. Front. ICT **5**, 1 (2018)
11. Bao, Y., Sun, F., Hua, X.: Operation action recognition using wearable devices with inertial sensors. In: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. IEEE (2017)
12. Karungaru, S.: Human action recognition using wearable sensors and neural networks. In: Control Conference, pp. 1–4. IEEE (2015)
13. Zhang, T., Wang, J., Xu, L.: Fall detection by wearable sensor and one-class SVM algorithm. In: Huang, D.S., Li, K., Irwin, G.W. (eds.) Intelligent Computing in Signal Processing and Pattern Recognition. LNCIS, vol. 345, pp. 858–863. Springer, Berlin (2016). https://doi.org/10.1007/978-3-540-37258-5_104
14. Thomas, O., Sunehag, P., Dror, G.: Wearable sensor activity analysis using semi-Markov models with a grammar. Pervasive Mob. Comput. **6**(3), 342–350 (2010)
15. Gjoreski, H., Lustrek, M., Gams, M.: Accelerometer placement for posture recognition and fall detection. In: Seventh International Conference on Intelligent Environments, pp. 47–54. IEEE Computer Society (2011)
16. Mannini, A., Sabatini, A.M., Intille, S.S.: Accelerometry-based recognition of the placement sites of a wearable sensor. Pervasive Mob. Comput. **21**, 62–74 (2015)
17. Stelios, M.A., Nick, A.D., Effie, M.T., Dimitris, K.M., Thomopoulos, S.C.A.: An indoor localization platform for ambient assisted living using UWB. In: International Conference on Advances in Mobile Computing and Multimedia, pp. 178–182 (2008)