

Chapter 1

Pervasive Sensing



Yiqiang Chen

Abstract The development of chips, sensors, and tele-communication, etc., with integrated sensing brings more opportunities to monitor various aspects of personal behavior and context. Especially, with the widespread use of intelligent devices and smart home infrastructure, it is more possible and convenient to sense users' daily life. Two common information of daily life is location and activity. Location information can reveal the places of important events. Activity information can expose users' health conditions. Besides these two kinds of information, other context also can be useful for assisting living. Hence, in this chapter, we will introduce some state-of-the-art user context sensing techniques under smart home infrastructure, including accurate indoor localization, fine-grained activity recognition, and pervasive context sensing. With the continuous sensing of location, activity, and other contextual information, it is possible to discovery users' life patterns which are crucial for health monitoring, therapy, and other services. What is more, it will bring more opportunities for improving the quality of peoples' life.

Keywords Pervasive sensing · Indoor localization · Activity recognition · Context sensing

1.1 Accurate Indoor Localization

Do you know how to accurately get you location information under unpredictable changes in environmental conditions? In recent years, with the development of mobile Internet, location-based services (LBSs) [1] have been widely used in our daily life, expanded from traditional navigation to real-time applications such as shared mobility and social network. With the development of LBS applications, the location area extends from outdoors to indoors, which creates great requirement of indoor localization with high accuracy. Indoor localization can be implemented in a variety of ways, such as base station, video, infrared, Bluetooth, Wi-Fi [2]. In which, Wi-Fi-based indoor localization has become the most popular way because of the wide

Y. Chen (✉)

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
e-mail: yqchen@ict.ac.cn

© Springer Nature Switzerland AG 2020

F. Chen et al. (eds.), *Smart Assisted Living*, Computer Communications and Networks, https://doi.org/10.1007/978-3-030-25590-9_1

coverage of Wi-Fi access points and the rapid development of intelligent terminals [3–5]. Although the research of indoor localization based on Wi-Fi has made great progress, in highly dynamic environments, due to the influence of multipath effect, environment changing and personnel flows, the fluctuation of wireless signal is large. High accuracy indoor localization still faces the problems of (1) the lack of large-scale labeled data in data layer, (2) the fluctuation of signal strength in feature layer, and (3) the weak adaption ability in model layer, which resulting in low location accuracy, rough trajectory granularity, and weak robustness. For the challenges above, this section will introduce some accurate indoor localization techniques.

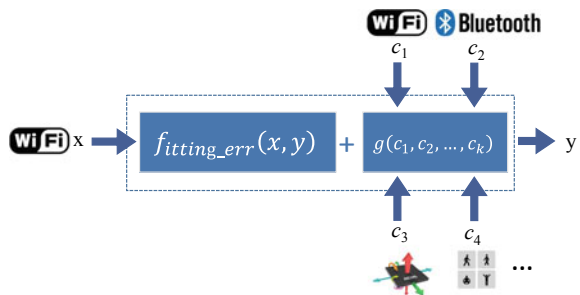
1.1.1 Context-Adaptive Localization Model

The wireless signal fingerprint-based indoor localization model is actually a mapping between the high-dimensional signal space and the physical space. For this kind of mapping models, the input \mathbf{x} is the feature vector extracted from the wireless signal strength, and the output \mathbf{y} is the position coordinate. Training the location model is to optimize the objective function $f = \operatorname{argmin}_f \sum_{i=1}^N |f(x_i) - y_i|^2$ with the given samples $\{(x_i, y_i) | i = 1, \dots, N\}$.

However, for highly dynamic environments, a context-adaptive model is necessary. This adaptive model should include the minimization of fitting errors and the context-adaptive constraints, as shown in Fig. 1.1, where $f_{\text{fitting_err}}(x, y)$ represents the fitting errors between model's output and calibration results, and $g(c_1, c_2, \dots, c_k)$ represents the constraints constructed with multi-source information of c_1, c_2, \dots, c_k . In addition, it is flexible to construct these constraints' information according to specific scenarios context, including multi-source signals, motion information, and user activities.

Compared with existing methods, the model has three advantages: (1) It gives a unified optimization objective, providing a reference for constructing multi-source information fusion localization method; (2) it realizes multi-source information fusion on the model level, fully mining the correlation and redundancy between

Fig. 1.1 Context-adaptive location model for high dynamic environment



multi-source information; (3) it has more flexible constraints, making the model scalable for any kind of high dynamic environments.

1.1.2 *Semi-supervised Localization Model with Signals' Fusion*

Aiming at the problem of low location accuracy caused by the lack of large-scale labeled data, a semi-supervised localization model based on multi-source signals fusion is introduced here. This model combines the fitting error term of the labeled data and manifold constraint terms of the Wi-Fi and Bluetooth signals and optimizes the objective equation by adjusting the weight coefficient of all terms. The experimental results [6] showed that the method based on multi-source signals fusion can achieve optimal location results when applied to the location problem of sparse calibration, and the location accuracy was higher than that of the existing supervised learning methods and semi-supervised learning methods.

Unlike previous single-signal-based semi-supervised manifold methods [7–12], it is better to combine the Wi-Fi and BLE signals into a single model. To the best of our knowledge, Wi-Fi and BLE signals have different propagation characteristics and effective distances. When considering both of Wi-Fi and BLE in a semi-supervised learning model, it should separately build the manifold regularization for each of them.

In accordance with the structural risk minimization principle [13], FSELM [6] used graph Laplacian regularization to find the structural relationships of both the labeled and unlabeled samples in the high-dimensional signal space. For the construction of a semi-labeled graph G based on l labeled samples and u unlabeled samples, each collected signal vector $s_j = [s_{j1}, s_{j2}, \dots, s_{jN}] \in R^N$ is represented by a vertex j , and if the vertex j is one of the neighbors of i , represented by drawing an edge with a weight of w_{ij} connecting them. According to Belkin et al. [14], the graph Laplacian L can be expressed as $L = D - W$. Here, $W = [w_{ij}]_{(l+u) \times (l+u)}$ is the weight matrix, where $w_{ij} = \exp(-\|s_i - s_j\|^2 / 2\sigma^2)$ if s_i and s_j are neighbors along the manifold and $w_{ij} = 0$ otherwise, and D is a diagonal matrix given by $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$. As illustrated in Fig. 1.2, to consider the empirical risk while controlling the complexity, FSELM minimized the fitting error plus two separate smoothness penalties for Wi-Fi and BLE as (1.1):

$$\operatorname{argmin}_f \left\{ \frac{1}{2} \|\mathbf{f} - \mathbf{T}\|^2 + \lambda_1 \mathbf{f}^T \mathbf{L}_1 \mathbf{f} + \lambda_2 \mathbf{f}^T \mathbf{L}_2 \mathbf{f} \right\} \quad (1.1)$$

The first term represents the empirical error with respect to the labeled training samples. The second and third terms represent the manifold constraints for Wi-Fi and BLE based on the graph Laplacians L_1 and L_2 . By adjusting the two coefficients

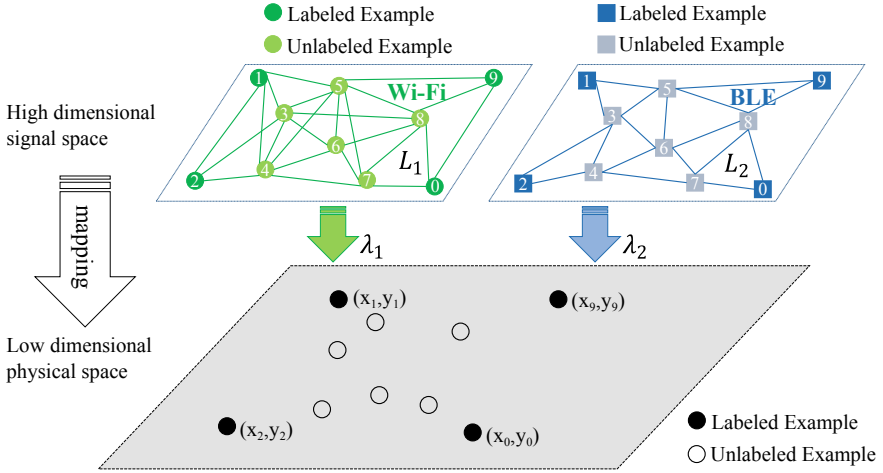


Fig. 1.2 FSELM model illustration (L_1 and L_2 are the graph Laplacians of the Wi-Fi and BLE signals, and λ_1 and λ_2 are the weight coefficients of the two manifold constraints)

λ_1 and λ_2 , it can control the relative influences of the Wi-Fi and BLE signals on the model.

When applied to sparsely calibrated localization problems, FSELM is advantageous in three aspects. Firstly, it dramatically reduces the human calibration effort required when using a semi-supervised learning framework. Secondly, it uses fused Wi-Fi and BLE fingerprints to markedly improve the location accuracy. Thirdly, it inherits the beneficial properties of ELMs in terms of training and testing speed because the input weights and biases of hidden nodes can be generated randomly. The findings indicate that effective multi-data fusion can be achieved not only through data layer fusion, feature layer fusion, and decision layer fusion but also through the fusion of constraints within a model. In addition, for semi-supervised learning problems, it is necessary to combine the advantages of different types of data by optimizing the model's parameters. These two contributions will be valuable for solving other similar problems in the future.

1.1.3 Motion Information Constrained Localization Model

For Wi-Fi fingerprint-based indoor localization, the basic approach is to fingerprint locations of interest with vectors of RSS of the access points during offline phase and then locate mobile devices by matching the observed RSS readings against this database during online phase. By this way, continuous localization can be summarized as trying to find a smooth trajectory going through all labeled points. Thus, in order to recover the trajectory, it still needs a certain number of labeled data, especially in some important positions (e.g., corners).

Considering that a user holds a mobile phone and walks in an indoor wireless environment with n Wi-Fi access points inside. At some time t , the signal received from all the n access points is measured by the mobile devices to form a signal vector $s_t = [s_{t1}, s_{t2}, \dots, s_{tn}] \in R^n$. As time goes on, the signal vectors will come in stream manner. After a period of time, a sequence of m vectors will be obtained from mobile phone and form a $m \times n$ matrix $S = [s_1^T, s_2^T, \dots, s_m^T]$, where the ‘T’ indicates matrix transposition. Along the user’s trajectory, only some places are known and labeled, and the rest are unknown. The purpose is to generate and update the trajectory points which can form a $m \times 2$ matrix $P = [p_1^T, p_2^T, \dots, p_m^T]$, where $p_t = [x_t, y_t]^T$ is the location of mobile device at time t . Meanwhile, for each step, the user heading orientation can also be obtained from mobile devices in every time t . Thus, while collecting the RSS, another vector of m orientation values can be generated: $O = [o_1, \dots, o_t, \dots, o_m]^T$. Here, o_t indicates the angle to north in horizontal plane, which is called azimuth. With the Wi-Fi signal matrix and the orientation vector, the mapping function should be $f(S, O) = P$. By this way, it can supplement the location for these unlabeled data, reducing the calibration work.

The fusion mapping model $f(S, O) = P$ from the signal space to the physical space can be optimized by $f^* = \operatorname{argmin}_f \sum_{i=1}^l |f_i - y_i|^2 + \delta \sum_{i=1}^l |o_{f_i} - o_{y_i}|^2 + \gamma f^T L f$, where the first term measures the fitting error to the labeled points, the second term is the fitting error to the user heading orientation offered by mobile phone, and the third term refers to the manifold graph Laplacian.

It brings good performance for both tracking mobile nodes and manual calibration reduction in wireless sensor networks. This model is based on two observations: (1) similar signals from access points imply close locations; (2) both labeled data positions and the real-time orientations can help tracking the traces. Thus, it learned a mapping function between the signal space and the physical space conjoin a few labeled data and a large amount of unlabeled data, and the constraint of orientation obtained from mobile devices.

The experimental results [15] showed that this method can achieve a higher tracking accuracy with much less calibration effort. It is robust to reduce the number of calibrated data. Furthermore, if applying it for offline calibration, the online location is much better than some other methods before. Moreover, it can reduce time consumption by parallel processing while maintaining trajectory learning accuracy.

1.2 Fine-Grained Activity Recognition

Traditional activity recognition methods aim at discovering pre-defined activity with body-attached sensors such as accelerometers and gyroscopes. However, peoples’ activities are so diverse; they cannot be covered by some pre-defined activities. As the way the devices are worn, the location the devices are placed, the person who wears the devices, etc., which all lead to the decreasing the recognition accuracy. And it needs a large amount of labeled data to maintain the recognition performance.

In this section, we will show the methods including transfer learning, generative adversarial networks (GANs), incremental learning to implement fine-grained activity recognition with less human labor.

1.2.1 Transfer Learning-Based Activity Recognition

The combination of sensor signals from different body positions can be used to reflect meaningful knowledge such as a person's detailed health conditions [16] and working states [17]. However, it is nontrivial to design wearing styles for a wearable device. On the one hand, it is not comfortable to equip all the body positions with sensors which make the activities restricted. Therefore, we can only attach sensors on limited body positions. On the other hand, it is impossible to perform HAR if the labels on some body parts are missing, since the activity patterns on specific body positions are significant to capture certain information.

Assume a person is suffering from small vessel disease (SVD) [18], which is a severe brain disease heavily related to activities. However, we cannot equip his all body with sensors to acquire the labels since this will make his activities unnatural. We can only label the activities on certain body parts in reality. If the doctor wants to see his activity information on the arm (we call it the target domain), which only contains sensor readings instead of labels, how to utilize the information on other parts (such as torso or leg, we call them the source domains) to help obtain the labels on the target domain? This is referred to as the cross-position activity recognition (CPAR).

To tackle the above challenge, several transfer learning methods have been proposed [19]. The key is to learn and reduce the distribution divergence (distance) between two domains. With the distance, we can perform source domain selection as well as knowledge transfer. Based on this principle, existing methods can be summarized into two categories: exploiting the correlations between features [20, 21], or transforming both the source and the target domains into a new shared feature space [22–24].

Existing approaches tend to reduce the global distance by projecting all samples in both domains into a single subspace. However, they fail to consider the local property within classes [25]. The global distance may result in loss of domain local property such as the source label information and the similarities within the same class. Therefore, it will generate a negative impact on the source selection as well as the transfer learning process. It is necessary to exploit the local property of classes to overcome the limitation of global distance learning.

This chapter introduces a Stratified Transfer Learning (STL) framework [26] to tackle the challenges of both source domain selection and knowledge transfer in CPAR. The term 'stratified' comes from the notion of splitting at different levels and then combining. The well-established assumption that data samples within the same class should lay on the same subspace, even if they come from different domains [27] is adopted. Thus, 'stratified' refers to the procedure of transforming features

into distinct subspaces. This has motivated the concept of stratified distance (SD) in comparison to traditional global distance (GD). STL has four steps:

1. **Majority Voting:** STL uses the majority voting technique to exploit the knowledge from the crowd [28]. The idea is that one certain classifier may be less reliable, so we assemble several different classifiers to obtain more reliable pseudo labels. To this end, STL makes use of some base classifiers learned from the source domain to collaboratively learn the labels for the target domain.
2. **Intra-class Transfer:** In this step, STL exploits the local property of domains to further transform each class of the source and target domains into the same subspace. Since the properties within each class are more similar, the intra-class transfer technique will guarantee that the transformed domains have the minimal distance. Initially, source domain and target domain are divided into C groups according to their (pseudo) labels, where C is the total number of classes. Then, feature transformation is performed within each class of both domains. Finally, the results of distinct subspaces are merged.
3. **Stratified Domain Selection:** A greedy technique is adopted in STL-SDS. We know that the most similar body part to the target is the one with the most similar structure and body functions. Therefore, STL uses the distance to reflect their similarity. It calculates the stratified distance between each source domain and the target domain and selects the one with the minimal distance.
4. **Stratified Activity Transfer:** After source domain selection, the most similar body part to the target domain can be obtained. The next step is to design an accurate transfer learning algorithm to perform activity transfer. This chapter introduces a Stratified Activity Transfer (STL-SAT) method for activity recognition. STL-SAT is also based on our stratified distance, and it can simultaneously transform the individual classes of the source and target domains into the same subspaces by exploiting the local property of domains. After feature learning, STL can learn the labels for the candidates. Finally, STL-SAT will perform a second annotation to obtain the labels for the residuals.

1.2.2 *GAN-Based Activity Recognition*

Transfer learning methods are effective ways to label practical unknown data, but they are incapable of generating realistic data. But fortunately, GANs framework is an effective way to generate labeled data from random noise space.

The vanilla GANs framework was firstly proposed in 2014 by Goodfellow et al. [29]. Since the GANs framework was proposed, it has been widely researched in many fields, such as image generation [29], image inpainting [30], image translation [31], super-resolution [32], image de-occlusion [33], natural language generation [34], text generation [35]. In particular, a great many variants of GANs have been widely explored to generate images with high fidelity, such as NVIDIA's progressive GAN [36], Google Deep Mind' BigGAN [37]. These variants of GANs provide

powerful methods for training resultful generative models that could output very convincing verisimilar images.

The original GANs framework is composed by a generative multilayer perceptron network and a corresponding discriminative multilayer perceptron network. The final goal of GANs is to estimate an optimal generator that can capture the distribution of real data with the adversarial assistance of a paired discriminator based on min-max game theory. The discriminator is optimized to differentiate the data distribution between authentic samples and spurious samples from its mutualistic generator. The generator and the discriminator are trained adversarially to achieve their optimization.

The optimization problem of the generator can be achieved by solving the formulation stated in 1.2:

$$\min_G V_G(D, G) = \min_G (\mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1.2)$$

The optimization problem of the discriminator can be achieved by solving the formulation stated in 1.3:

$$\max_D V_D(D, G) = \max_D (\mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1.3)$$

The final value function of the min-max game between the generator and the discriminator can be formulated as 1.4:

$$\min_G \max_D (D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1.4)$$

Firstly, the original GANs framework was proposed to generate plausible fake images approximating real images in low resolution, such as MNIST, TFD, CIFAR-10. Many straightforward extensions of GANs have demonstrated and leded one of the most potential research directions. Though the researches of GANs have gained great success in the field of generating realistic-looking images, the GANs framework has not been widely exploited for generating sensor data.

Inspired by the thought of GANs, Alzantot et al. [38] firstly tried idea of GANs to train the LSTM-based generator to produce sensor data, but their SenseGen is half-baked GANs' framework. Both the generator and the discriminator in SenseGen are trained separately; that is, the training process of the generator in SenseGen is not based on the back-propagating gradient from the discriminator.

In order to improve the performance of human activity recognition when a small number of sensor data are available under some special practical scenarios and resource-limited environments, Wang et al. [39] proposed SensoryGANs models. To the best of our knowledge, SensoryGANs models are the first unbroken generative adversarial networks applied in generating sensor data in the HAR research field. The specific GANs models were designed for three human daily activities, respectively. The generators accept the Gaussian random noises and output accelerometer data of the target human activity. The discriminators accept both the real accelerometer

sensor data and the spurious accelerometer sensor data from the generators and then output the probability of whether the input samples are from the real distribution. With the improvement of SensoryGANs, the research of human activity recognition, especially in resource-constrained environments, will be greatly encouraged.

Then, Yao et al. [40] proposed SenseGAN to leverage the abundant unlabeled sensing data, thereby minimizing the need for labeling effort. SenseGAN jointly trains three components, the generator, the discriminator, and a classifier. The adversarial game among the three modules can achieve their optimal performance. The generator receives random noises and labels and then outputs spurious sensing data. The classifier accepts sensing data and outputs labels. The samples from the classifier and the generator are both fed to the discriminator for differentiating the joint data/label distribution between real sensing data and spurious sensing data. Compared with supervised counterparts as well as other supervised and semi-supervised baselines, SenseGAN achieves substantial improvements in accuracy and F1 score. With only 10% of the originally labeled data, SenseGAN can attain nearly the same accuracy as a deep learning classifier trained on the fully labeled dataset.

1.2.3 Incremental Learning-Based Activity Recognition

With more labeled data, it becomes possible to get fine-grained activity. However, traditional sensor-based activity recognition methods train fixed classification models with labeled data collected off-line, which are unable to adapt to dynamic changes in real applications. With the emergence of new wearable devices, more diverse sensors can be used to improve the performance of activity recognition. While it is difficult to integrate a new sensor into a pre-trained activity recognition model, the emergence of new sensors will lead to a corresponding increase in the feature dimensionality of the input data, which may result in the failure of a pre-trained activity recognition model. The pre-trained activity recognition model is unable to take advantage of this new source of data.

To take advantage of data generated by new sensors, feature incremental learning method is an effective method. To improve the performance of indoor localization with more sensors, Jiang et al. [41] proposed a novel feature incremental and decremental learning method, namely FA-OSELM. It is able to adapt to the dynamic changes of sensors flexibly. However, the performance of FA-OSELM fluctuates heavily. Hou and Zhou [42] proposed the One-Pass Incremental and Decremental learning approach (OPID), which is able to adapt to evolving features and instances simultaneously. Xing et al. [43] proposed a perception evolution network that integrates the new sensor readings into the learned model. However, the impact of the sensor order is not considered.

Hu et al. [44] proposed a novel feature incremental activity recognition method, which is named Feature Incremental Random Forest (FIRF). It is able to adapt an existing activity recognition model to newly available sensors in a dynamic environment. Figure 1.3 shows an overview of the method.

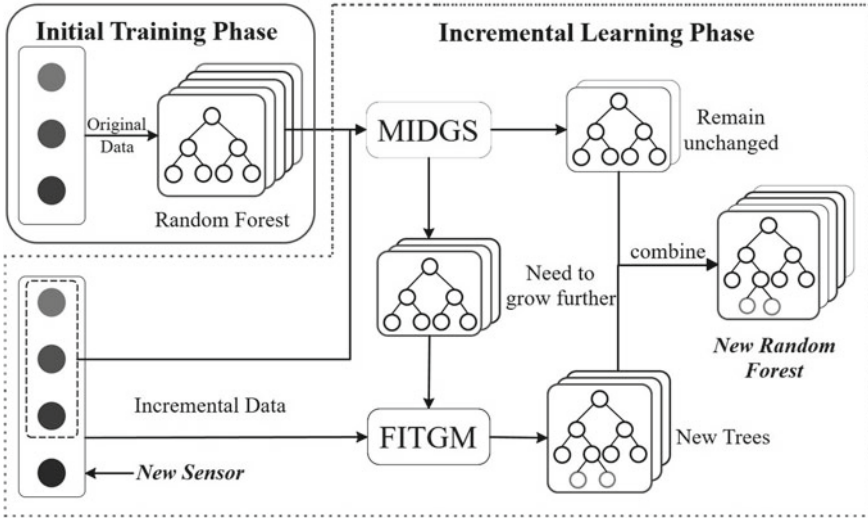


Fig. 1.3 Feature incremental random forest algorithm

In FIRE, there are two new strategies: (1) MIDGS which encourages diversity among individual decision trees in the incremental learning phase by identifying the individual learners that have high redundancy with the other individual learners and low recognition accuracy, and (2) FITGM which improve the performance of these identified individual decision trees with new data collected from both existing and newly emerging sensors.

In real applications, people may learn new motion activities over time, which is usually classified as dynamic changes in class. When a new kind of activity is performed or the behavioral pattern changes over time, devices with preinstalled activity recognition models may fail to recognize new activities or even known activities with changed manners. To adapt to the changes of activities, traditional batch learning methods require retraining the whole model from scratch. This will result in a great waste of time and memory.

Class incremental learning method is an effective way to address this problem. Different from batch learning, incremental learning, or online learning methods update existing models with new knowledge. In [45], Zhao et al. presented a class incremental extreme learning machine (CIELM), which adds new output nodes to accommodate new class data. With update to output weights, CIELM can recognize new activities dynamically. Camoriano et al. [46] employed recursive technique and regularized least squares for classification (RLSC) to seamlessly add new classes to the learned model. They considered the imbalance between classes in the class incremental learning phase. Zhu et al. [47] introduced a framework, namely the one-pass class incremental learning (OPCIL), to handle new emerging classes. They proposed a pseudo instances generating approach to address the new class adaptation issue. Ristin et al. [48] put forward two variants of random forest to incorporate new classes.

Four incremental learning strategies are devised to exploit hierarchical nature of random forest for efficient updating.

In [49], Hu et al. proposed an effective class incremental learning method, named class incremental random forest (CIRF), to enable existing activity recognition models to identify new activities. They designed a separating axis theorem-based splitting strategy to insert internal nodes and adopt Gini index or information gain to split leaves of the decision tree in the random forests. With these two strategies, both inserting new nodes and splitting leaves are allowed in the incremental learning phase. They evaluated their method on three UCI public activity datasets and compared with other state-of-the-art methods. Experimental results show that their incremental learning method converges to the performance of batch learning methods (random forests and extremely randomized trees). Compared with other state-of-the-art methods, it is able to recognize new class data continuously with a better performance.

1.3 Pervasive Context Sensing

With the pervasiveness of intelligent hardware, more individual context can be sensed, which is meaningful to infer users' life patterns, health conditions, etc. In this section, we will introduce context sensing methods with pervasive intelligent hardware, including sleep sensing, household water-usage sensing, etc.

1.3.1 *Sleeping Sensing*

Sleeping is a vital activity that people spend nearly a third of lifetime to do. Many studies have shown that sleep disorder is related to many serious diseases including senile dementia, obesity, and cardiovascular disease [50]. Clinical studies have reported that sleeping is composed of two stages including rapid eye movement (REM) and non-rapid eye movements (NREM). NREM can be further divided into light and deep sleep stages. During sleep, REM and NREM change alternately. Babies can spend up to 50% of their sleep in the REM stage, compared to only about 20% for adults. As people getting older, they sleep more lightly and get less deep sleep. Therefore, it is meaningful to find out the distribution of different sleep stages.

As sleep quality is very important for health, a lot of previous researches have been done on sleep detection. The methods of analyzing sleep quality mainly monitor different sleep stages. Recently, the technologies of recording sleep stages are divided into two categories. One category is polysomnography (PSG)-based approaches [51]. PSG monitors many body functions including brain (EEG), eye movements (EOG), skeletal muscle activation (EMG), and heart rhythm (ECG) during sleep. However, collecting the polysomnography signals or brain waves requires professional equipments and specialized knowledge. Another category is actigraphy-based approaches. Typical devices can be divided into the following two categories. The first category

is wearable sleep and fitness tracker such as Fitbit charge 2 and Jawbone Up [52]. These devices primarily work by actigraphy. Several algorithms [53] utilized wrist activity data to predict sleep/wake states. The results have shown that the accuracy of predicting sleep/wake through recording wrist activity data approaches score using EEG data. But wearable sleep devices have some weaknesses because of accuracy concerns for sleep stages. These devices detect sleep stages based on logged acceleration data generated by body movement. This means if a user does not move, these devices have to rely on other auxiliary sensors. The second category is non-wearable sleep trackers such as Beddit 3.0 Smart Sleep Monitor. These are dedicated sleep trackers that users do not wear on wrist. They tend to provide more detailed sleep data. Many products use non-wearable smartphone sensors to assess sleep quality or sleep stage. An application called iSleep [54] used the microphone of smartphone to detect the sleep events. The method extracts three features to classify different events including body movement, snoring, and coughing. These non-wearable sleep trackers tend to use many sensors on smartphone and a lot of manual intervention to extract features.

Different from these works, the work [55] leveraged microphone without any other auxiliary sensor or much manual intervention to detect sleep stages. Acoustic signal collected by the microphone is sensitive enough to record information. After the acoustic signal is collected, the spectrogram visual representation is given. Specifically, the spectrogram is the magnitude squared of the short-time Fourier transform (STFT). It splits time signal to short segments of equal length and then computes STFT on each segment.

Once the spectrogram has been computed, they can be processed by the deep learning model. Deep learning is a new area of machine learning research. Its algorithms build a large set of layers to extract a hierarchy of features from low level to high level. Deep learning models include deep neural network (DNN), convolutional neural network (CNN, or ConvNet), etc. ConvNet [56] is the most efficient approach for image and speech recognition. The major difference between ConvNet and ordinary neural networks is that ConvNet architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture and vastly reduce the amount of parameters in the network.

The convolutional neural network architecture and training procedure are shown in Fig. 1.4. Learning from the relatively good effect of the network configuration in image recognition, this configuration can improve the expression ability of ConvNet. At the same time, accumulating convolutional layers and pooling layers guarantees long-range dependence (LRD) of acoustic signal, which is more robust than conventional ConvNet architecture.

During the training process, the goal is to minimize the loss function in backward propagation. The optimizers such as stochastic gradient descent (SGD) and Nadam are used to update the weights of hidden layers. The output of network is divided into three categories, deep sleep, light sleep, and REM.

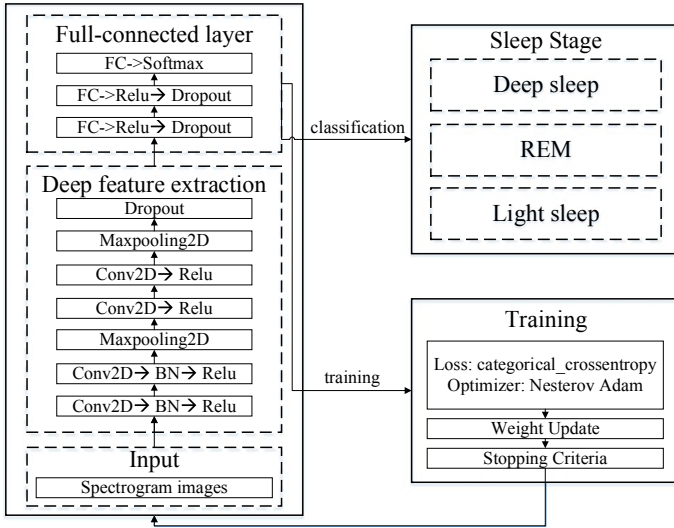


Fig. 1.4 Illustration of the convolutional neural networks (CNNs) architecture and training procedure

1.3.2 Water-Usage Sensing

A person’s daily activities can be recognized by monitoring the infrastructures (e.g., water, electric, heat, ventilating, air conditioning) in the house. Infrastructure-mediated sensing has been recognized as a low-cost and nonintrusive activity recognition technique.

Several infrastructure-mediated sensing approaches for water-usage activity recognition have been proposed recently. A water-usage activity recognition technique was proposed by Fogarty [57], deploying four microphones on the surface of water pipes near the inlets and outlets. Froehlich et al. [58] proposed HydroSense, another infrastructure-mediated single-point sensing technique. Thomaz et al. [59] proposed an learning approach for high-level activity recognition, which combined single-point, infrastructure-mediated sensing with a vector space model. Their work has been considered to be the first one of employing the method for inferring high-level water-usage activities. However, the infrastructure of the house has to be remodeled in order to work in with the installation of the pressure sensors.

To solve the above question, a single-point infrastructure-mediated sensing system for water-usage activity recognition proved to be effective which has a single 3-axis accelerometer clinging to the surface of the main water pipe in the house [60]. The structure of the water pipe in the apartment can be seen in Fig. 1.5. The thick and thin gray lines represent the main water pipe and branches of the main pipe. The green circle and red star in Fig. 1.5 are water meter and the accelerometer, respectively.

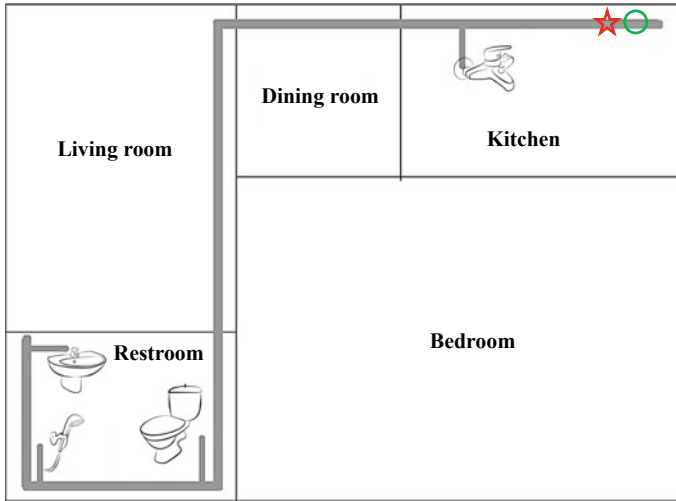


Fig. 1.5 Structure of the water pipe

The water-usage activity recognition system has six modules, which are:

A. Data Preprocessing

Normally, there exist some noises in the raw time series samples which should be filtered out. The median filter technique is employed in this data preprocessing module, and the filter window is set to 3.

B. Segmentation

The segmentation module is aimed at segmenting both the rugged segments (time series rugged samples) and smooth segments (time series smooth samples) from time series samples.

First, sample windows are generated on the set of time series samples according to the sliding window mechanism; second, annotate each sample window to be rugged or smooth based on whether its standard deviation is no less than a threshold or not. At last, a rugged (or smooth) segment is defined as a time series rugged (or smooth) windows.

C. Data Post-processing

The data post-processing module is to make all the rugged segments generated in the previous module completer and more precise.

First-stage post-process procedure: In the first circumstance, any smooth segment (in between two rugged segments), whose corresponding samples are no more than a threshold, is re-annotated to be rugged segments. After that, all the neighboring rugged segments make up a long-rugged segment.

Second-stage post-process procedure: In the second one, any rugged segment (in between two smooth segments), whose corresponding samples are no more than another threshold, is re-annotated to be smooth segments. After that, all the neighboring smooth segments make up a long smooth segment.

D. *Feature Extraction*

Instances are generated by utilizing the sliding window mechanism again on each rugged segment. The feature extraction module is executed on each sub-segment. Eight features (0.25-quantile, 0.5-quantile, 0.75-quantile, mean value, standard deviation, quadratic sum, zero-crossing, spectral peak) are extracted from a window of sample values in each axis (x -axis, y -axis, or z -axis in the accelerometer device). In all, there are 24 features for each instance.

E. *Model Generation and Prediction*

All the instances are split into two sets (the training set and the testing set) with approximately the same size. Instances in the same segment are assured to put into the same set, since you do not want any water-usage activity to be apart.

Support vector machine (SVM) is employed for model generation, and Gaussian kernel can be utilized as its kernel function. Two parameters—the kernel parameter and the penalty parameter—need to be set before starting the learning process. In the end, a classifier is constructed on the training set.

The classifier is then employed to predict the labels of instances in the testing set (testing instances). These prediction results are recognized as SVM's prediction labels for the testing instances.

F. *Prediction Results' Fusion*

The prediction results' fusion module is done by law of 'The minority is subordinate to the majority'. Specifically, for each water-usage activity, the number of testing corresponds to the most testing instances. In the end, the prediction labels of all instances in the segment are replaced by the dominant water-usage activity. The prediction results of the rugged segment are fused finally.

The nonintrusive and single-point infrastructure-mediated sensing approach in this chapter can recognize 4-class water-usage activities in daily life. Data is collected unobtrusively by a single low-cost 3-axis accelerometer attached to the surface of the main water pipe in the house, making the installation process much more convenient.

1.3.3 *Non-contact Physiological Signal Sensing*

Non-contact vital sign detection has received significant attention from healthcare researchers, for it can perform basic physiological signal acquisition without any interference to the user. The electrode-attached ones, such as electrocardiography (ECG) or respiratory detection instrument, need the user fixed in a particular place or to be worn by the user all day long. These approaches have a negative impact on the daily life of users, which cannot be used in many applications, such as sleep apnea monitoring, burned patients' vital sign monitoring, and health care jobs that require long-term monitoring.

The heartbeats and respiratory are common physiological signals, which can be used for sleep monitoring and abnormal body monitoring. At present, the traditional

methods for heartbeat detection are electrocardiogram (ECG) and photoplethysmography (PPG). The traditional detection method for breathing is mainly measuring the air volume and flow of the nose and mouth through the breathing process. All these methods require direct physical contact with the user, and the electrodes, sensors, masks need to be placed close to the skin for physiological signal measurement. Although the measurement result is more accurate, it has a strong interference to the normal life of the user, greatly reduces the comfort of the user, and cannot achieve long-term monitoring of the physiological information of the user. Therefore, non-contact detection methods attract more interest recently.

The non-contact detection of heartbeat and respiratory can be achieved by many methods, such as camera [61], radar, Wi-Fi, ultrasonic [62]. The camera method is to perform heartbeat detection through face video and perform respiratory detection by using body video. The other methods mainly perform heartbeat and respiratory frequency detection by detecting chest vibration caused by respiratory and heartbeat. Among them, the radar method has better recognition effect when the user is still, for electromagnetic can penetrate the clothes or covers and most of it will be reflected when it reaches the surface of the human body.

The radar method also can be subdivided according to the principle of signal transmission and reception. The most used radar methods are Doppler radar, FMCW radar [63]. Also, there are many innovative radar methods are used in heartbeat and respiratory detection, such as UWB pulse radar [64], self-injection-locked radar [65], UWB impulse radar [66].

The Doppler Radar: The Doppler radar method measures a user's chest movement via the return signal phase. Doppler radar transmits continue wave (CW) electromagnetic wave toward the user's body, and the RF signal will be reflected from the skin and tissue of the body. The receiver acquires the electromagnetic signal and mixes the received signal with the transmitter signal for vital signal detection.

Recently, the coherent receiver is used by mixing the received signal with a quadrature mixer. The quadrature mixer mixes the original received signal and a 90-degree shifted signal with the transmitter signal to achieve two quadrature components. With this method, the NULL point of radar detection is avoided. The signal needs to be demodulation with linear or nonlinear demodulation methods to get the phase change containing $x(t)$. Then the heartbeat and the respiratory signal can be achieved with signal processing methods or machine learning methods.

The FMCW Radar: The frequency modulated continuous wave (FMCW) radar can determine the absolute distance between the system and a target. The FMCW radar transmits variable frequency signal with a modulation frequency being able to slew up and down as sine wave, sawtooth wave, triangle wave, or square wave [63]. And for vital sign detection, if the target is a man, the received signal will contain the information of the chest movement. Then, the signal will have a frequency shift of the chest motion frequency. By detecting the frequency shift in the range information, the heartbeat and respiratory frequency can be calculated.

1.4 Conclusions

In this chapter, we show different ways to sense user's location information, activity information, and other context information with the pervasiveness of intelligent devices under smart home infrastructure. In the future, with the development of Internet of things (IoT), edge computing, and cloud computing, the sensing ability in smart home will be unprecedentedly powerful. And the collaborative computing framework of the above three (IoT, edge computing and cloud computing) would be the trend, which can adaptively use the device and resource to optimally achieve the task, what is more, with the maturity of pervasive sensing techniques, it will bring more convenience to people's daily life and make high-quality living possible.

References

1. Prasad M (2002) Location based services. *GIS Dev* 3–35
2. Xiang Z, Song S, Chen J et al (2004) A wireless LAN-based indoor positioning technology. *IBM J Res Dev* 48(5.6):617–626
3. Liu H, Darabi H, Banerjee P, Liu J (2007) Survey of wireless indoor positioning techniques and systems. *IEEE Trans Syst Man Cybern Part C* 37(6):1067–1080
4. Kjægaard MB (2007) A taxonomy for radio location fingerprinting. In: Hightower J, Schiele B, Strang T (eds) *Location- and contextawareness*, LNCS, vol 4718. Springer, Berlin, pp 139–156
5. Brunato M, Battiti R (2005) Statistical learning theory for location fingerprinting in wireless LANs. *Comput Netw* 47:825–845
6. Jiang X, Chen Y, Liu J, Gu Y, Hu L (2018) FSELM: fusion semi-supervised extreme learning machine for indoor localization with Wi-Fi and Bluetooth fingerprints. *Soft Comput* 22(11):3621–3635
7. Liu J, Chen Y, Liu M, Zhao Z (2011) Selm: semi-supervised elm with application in sparse calibrated location estimation. *Neurocomputing* 74(16):2566–2572
8. Pan JJ, Yang Q, Chang H, Yeung D-Y (2006) A manifold regularization approach to calibration reduction for sensor-network based tracking. In: *AAAI*, pp 988–993
9. Pan JJ, Yang Q, Pan SJ (2007) Online co-localization in indoor wireless networks by dimension reduction. In: *Proceedings of the national conference on artificial intelligence*, vol 22. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, p 1102
10. Gu Y, Chen Y, Liu J, Jiang X (2015) Semi-supervised deep extreme learning machine for Wi-Fi based localization. *Neurocomputing* 166:282–293
11. Zhang Y, Zhi X (2010) Indoor positioning algorithm based on semisupervised learning. *Comput Eng* 36(17):277–279
12. Scardapane S, Comminiello D, Scarpiniti M, Uncini A (2016) A semisupervised random vector functional-link network based on the transductive framework. *Inf Sci* 364:156–166
13. Vapnik V (2013) *The nature of statistical learning theory*. Springer science & business media, Berlin
14. Belkin M, Niyogi P, Sindhvani V (2006) Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. *J Mach Learn Res* 7:2399–2434
15. Jiang X, Chen Y, Liu J, Gu Y, Hu L, Shen Z (2016) Heterogeneous data driven manifold regularization model for fingerprint calibration reduction. In: *2016 International IEEE conferences on ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress (UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld)*. IEEE, pp 74–81

16. Hammerla N Y, Fisher J, Andras P et al (2015) PD disease state assessment in naturalistic environments using deep learning. In: Twenty-Ninth AAAI conference on artificial intelligence
17. Plötz T, Hammerla NY, Olivier PL (2011) Feature learning for activity recognition in ubiquitous computing. In: Twenty-second international joint conference on artificial intelligence
18. Wardlaw J M, Smith EE, Biessels GJ et al (2013) Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurol* 12(8):822–838
19. Cook D, Feuz KD, Krishnan NC (2013) Transfer learning for activity recognition: a survey. *Knowl Inf Syst* 36(3):537–556
20. Blitzer J, McDonald R, Pereira F (2006) Domain adaptation with structural correspondence learning. In: Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 120–128
21. Kouw WM, Van Der Maaten LJP, Krijthe JH et al (2016) Feature-level domain adaptation. *J Mach Learn Res* 17(1):5943–5974
22. Pan SJ, Tsang IW, Kwok JT et al (2011) Domain adaptation via transfer component analysis. *IEEE Trans Neural Netw* 22(2):199–210
23. Gong B, Shi Y, Sha F et al (2012) Geodesic flow kernel for unsupervised domain adaptation. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 2066–2073
24. Long M, Wang J, Sun J et al (2015) Domain invariant transfer kernel learning. *IEEE Trans Knowl Data Eng* 27(6):1519–1532
25. Lin Y, Chen J, Cao Y et al (2017) Cross-domain recognition by identifying joint subspaces of source domain and target domain. *IEEE Trans Cybern* 47(4):1090–1101
26. Wang J, Chen Y, Hu L et al (2018) Stratified transfer learning for cross-domain activity recognition. In: 2018 IEEE international conference on pervasive computing and communications (PerCom). IEEE, pp 1–10
27. Elhamifar E, Vidal R (2013) Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans Pattern Anal Mach Intell* 35(11):2765–2781
28. Prelec D, Seung HS, McCoy J (2017) A solution to the single-question crowd wisdom problem. *Nature* 541(7638):532
29. Goodfellow I, Pouget-Abadie J, Mirza M et al (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680
30. Yeh RA, Chen C, Yian Lim T et al (2017) Semantic image inpainting with deep generative models. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5485–5493
31. Isola P, Zhu JY, Zhou T et al (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1125–1134
32. Ledig C, Theis L, Huszár F et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690
33. Zhao F, Feng J, Zhao J et al (2018) Robust lstm-autoencoders for face de-occlusion in the wild. *IEEE Trans Image Process* 27(2):778–790
34. Press O, Bar A, Bogin B et al (2017) Language generation with recurrent generative adversarial networks without pre-training. arXiv preprint [arXiv:1706.01399](https://arxiv.org/abs/1706.01399)
35. Yu L, Zhang W, Wang J et al (2017) Seqgan: sequence generative adversarial nets with policy gradient. In: Thirty-first AAAI conference on artificial intelligence
36. Karras T, Aila T, Laine S et al (2017) Progressive growing of gans for improved quality, stability, and variation. arXiv preprint [arXiv:1710.10196](https://arxiv.org/abs/1710.10196)
37. Brock A, Donahue J, Simonyan K (2018) Large scale gan training for high fidelity natural image synthesis. arXiv preprint [arXiv:1809.11096](https://arxiv.org/abs/1809.11096)
38. Alzantot M, Chakraborty S, Srivastava M (2017) Sensegen: a deep learning architecture for synthetic sensor data generation. In: 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops). IEEE, pp 188–193

39. Wang J, Chen Y, Gu Y et al (2018) SensoryGANs: an effective generative adversarial framework for sensor-based human activity recognition. In: 2018 international joint conference on neural networks (IJCNN). IEEE, pp 1–8
40. Yao S, Zhao Y, Shao H et al (2018) Sensegan: enabling deep learning for internet of things with a semi-supervised framework. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2(3):144
41. Jiang X, Liu J, Chen Y et al (2016) Feature adaptive online sequential extreme learning machine for lifelong indoor localization. *Neural Comput Appl* 27(1):215–225
42. Hou C, Zhou ZH (2018) One-pass learning with incremental and decremental features. *IEEE Trans Pattern Anal Mach Intell* 40(11):2776–2792
43. Xing Y, Shen F, Zhao J (2015) Perception evolution network adapting to the emergence of new sensory receptor. In: Twenty-fourth international joint conference on artificial intelligence
44. Hu C, Chen Y, Peng X et al (2018) A novel feature incremental learning method for sensor-based activity recognition. *IEEE Trans Knowl Data Eng*
45. Zhao Z, Chen Z, Chen Y et al (2014) A class incremental extreme learning machine for activity recognition. *Cognit Comput* 6(3):423–431
46. Camoriano R, Pasquale G, Ciliberto C et al (2017) Incremental robot learning of new objects with fixed update time. In: 2017 IEEE international conference on robotics and automation (ICRA). IEEE, pp 3207–3214
47. Zhu Y, Ting KM, Zhou ZH (2017) New class adaptation via instance generation in one-pass class incremental learning. In: 2017 IEEE international conference on data mining (ICDM). IEEE, pp 1207–1212
48. Ristin M, Guillaumin M, Gall J et al (2016) Incremental learning of random forests for large-scale image classification. *IEEE Trans Pattern Anal Mach Intell* 38(3):490–503
49. Hu C, Chen Y, Hu L et al (2018) A novel random forests based class incremental learning method for activity recognition. *Pattern Recognit* 78:277–290
50. Kanbay A, Buyukoglan H, Ozdogan N et al (2012) Obstructive sleep apnea syndrome is related to the progression of chronic kidney disease. *Int Urol Nephrol* 44(2):535–539
51. Berry RB, Brooks R, Gamaldo CE et al (2012) The AASM manual for the scoring of sleep and associated events. In: Rules, terminology and technical specifications. Darien, Illinois, American Academy of Sleep Medicine, p 176
52. Jawbone Up <https://jawbone.com/up>
53. Hoque E, Dickerson RF, Stankovic JA (2010) Monitoring body positions and movements during sleep using wisps. In: *Wireless health 2010*. ACM, pp 44–53
54. Hao T, Xing G, Zhou G (2013) iSleep: unobtrusive sleep quality monitoring using smartphones. In: *Proceedings of the 11th ACM conference on embedded networked sensor systems*. ACM, p 4
55. Zhang Y, Chen Y, Hu L et al (2017) An effective deep learning approach for unobtrusive sleep stage detection using microphone sensor. In: 2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI). IEEE Computer Society
56. LeCun Y, Bottou L, Bengio Y et al (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
57. Fogarty J, Au C, Hudson SE (2006) Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In: *Proceedings of the 19th annual ACM symposium on user interface software and technology*. ACM, pp 91–100
58. Froehlich JE, Larson E, Campbell T et al (2009) HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In: *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, pp 235–244
59. Thomaz E, Bettadapura V, Reyes G et al (2012) Recognizing water-based activities in the home through infrastructure-mediated sensing. In: *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, pp 85–94
60. Hu L, Chen Y, Wang S et al (2013) A nonintrusive and single-point infrastructure-mediated sensing approach for water-use activity recognition. In: 2013 IEEE 10th international conference on high performance computing and communications & 2013 IEEE international conference on embedded and ubiquitous computing. IEEE, pp 2120–2126

61. Qi H, Guo Z, Chen X et al (2017) Video-based human heart rate measurement using joint blind source separation. *Biomed Signal Process Control* 31:309–320
62. Min SD, Kim JK, Shin HS et al (2010) Noncontact respiration rate measurement system using an ultrasonic proximity sensor. *IEEE Sens J*, 10(11):1732–1739
63. Li C, Peng Z, Huang TY et al (2017) A review on recent progress of portable short-range noncontact microwave radar systems. *IEEE Trans Microw Theory Techn* 65(5):1692–1706
64. Hosseini SMAT, Amindavar H (2017) UWB radar signal processing in measurement of heart-beat features[C]. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1004–1007
65. Wang FK, Tang MC, Su SC et al (2016) Wrist pulse rate monitor using self-injection-locked radar technology. *Biosensors* 6(4):54
66. Cho HS, Park YJ, Lyu HK et al (2017) Novel heart rate detection method using UWB impulse radar. *J Signal Process Syst* 87(2):229–239