A Survey of Sensor Modalities for Human Activity Recognition

Bruce X. B. Yu[a], Yan Liu[b] and Keith C. C. Chan[c]
Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China

Keywords: IoT, Human Activity Recognition, Sensors.

Abstract: Human Activity Recognition (HAR) has been attempted by various sensor modalities like vision sensors, ambient sensors, and wearable sensors. These heterogeneous sensors are usually used independently to conduct HAR. However, there are few comprehensive studies in the previous literature that investigate the HAR capability of various sensors and examine the gap between the existing HAR methods and their potential application domains. To fill in such a research gap, this survey unfastens the motivation behind HAR and compares the capability of various sensors for HAR by presenting their corresponding datasets and main algorithmic status. To do so, we first introduce HAR sensors from three categories: vision, ambient and wearable by elaborating their available tools and representative benchmark datasets. Then we analyze the HAR capability of various sensors regarding the levels of activities that we defined for indicating the activity complexity or resolution. With a comprehensive understanding of the different sensors, we review HAR algorithms from perspectives of single modal to multimodal methods. According to the investigated algorithms, we direct the future research on multimodal HAR solutions. This survey provides a panorama view of HAR sensors, human activity characteristics and HAR algorithms, which will serve as a source of references for developing sensor-based HAR systems and applications.

1 INTRODUCTION

HAR research has been revitalized in recent years with plenty of emerging big data technologies that involve various Internet of Thing (IoT) sensors. HAR is a broad field of study concerned with identifying the specific movement or action of a person based on sensor data. HAR could have a plenty of application domains like healthcare, assisted living, surveillance and computational behavioral science, etc. For instance, the concept of cashier-free supermarket has been emerged in recent years since the announcement of Amazon Go and the release of its patents (Dilip, 2015; Gianna, 2015). Sensor technologies such as Radio Frequency IDentification (RFID), computer vision, and sensor fusion have been collectively attempted by companies for no-check stores to detect activities of customers, but none of those technologies has been popularized to mass production due to their high implementation cost and intrinsic limitations. Another important application domain of HAR is healthcare where Applied Behavioral Analysis (ABA) has been empirically verified as effective treatment or prevention therapy for Autism kids (Chouhan & Sharma, 2017). Besides, it could also be applied to early screening of dementia symptoms for the elderly to take effective prevention therapies (Petersen et al., 2001).

In these application domains, HAR requires varied levels of activity recognition resolutions to tackle their specific core problems. If activities in a grocery store or at a patient's home could be detected at a high resolution that supports advanced behavior understanding, we could believe that it will breed realistic benefits to our daily life. However, it remains lack of studies examining the characteristics of human activity and the required level of HAR for landing them to those application domains. The existing research of HAR usually focus on a single sensor modality such as inertial sensor (Avci et al., 2010; Bulling et al., 2014), vision sensor (Presti & La Cascia, 2016), or WSN (Alemdar & Ersoy, 2010). Although the use of single sensor modality has

[a] https://orcid.org/0000-0001-9905-8154
[b] https://orcid.org/0000-0003-4242-4840
[c] https://orcid.org/0000-0003-1709-2637

Yu, B., Liu, Y. and Chan, K.
A Survey of Sensor Modalities for Human Activity Recognition.
DOI: 10.5220/0010145202820294
In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020) - Volume 1: KDIR*, pages 282-294
ISBN: 978-989-758-474-9

achieved some progresses on a few public datasets, it might impede the potential progress of HAR area since the real-world data source of HAR is heterogeneously multimodal. To our best knowledge, the effective data fusion strategies for multimodal HAR solutions have not been thoroughly explored in previous literatures. To identify the potential research gaps, this survey disentangles HAR from three aspects: activities, sensors and algorithms.

This survey is one of the first attempts to examine the complexity levels of human activity and the state-of-the-art levels of HAR achieved by different sensors. Unlike most existing literatures of survey that focus on a specific sensor modality, this paper provides a systematic review of various IoT sensors and their HAR capabilities. Main contributions of this survey are threefold. First, it collectively compares advantages and disadvantages of different sensors, which can serve as a guidance to choose proper sensor modalities for developing applicable holistic real-world applications. Second, it provides a definition of activity levels and a categorization scheme of HAR that are used to analyze the HAR capability of different sensor modalities. Third, it reviews multimodal algorithms from both perspectives of traditional and Deep Learning (DL) methods, and direct the future research of HAR.

Regarding the structure of this survey, we start with introducing different sensors in Section 2. We then define the levels of human activity and analyze the HAR capability of different sensors by using the sample activities of a breakfast preparation activity routine in Section 3. In Section 4, we explore the future direction of HAR by reviewing data fusion and processing methods of both single modal and multimodal HARs. In Section 5, we draw a conclusion and direct the future work.

2 IoT SENSORS FOR HAR

Regarding IoT sensors used for HAR, some researches classify them into two rough categories: ambient sensors and wearable sensors (Acampora et al., 2013; Chen, Hoey, et al., 2012). Ambient sensors refer to sensors connected as a wireless mesh/dense network that monitor the whole indoor environment as well as human subjects. Wearable sensors are attached to clothing and body, or even implanted under the human skin. It is also common to classify HAR sensors to three categories: vision, ambient and wearable (Palumbo et al., 2014). Plenty of vision-based behavior analysis technologies and applications have immerged in recent years, among which depth sensors remarkably attracts the interest of researchers. Ambient sensors like RFID, pressure sensor and inferred have been attempted but not as popular as vision sensors. Wearable sensors like accelerometer and gyroscope are popularly adopted in both industrial and academic solutions. In the remaining of this section, we extensively introduce representative IoT sensors from three sensor categories: vision, ambient and wearable sensors.

2.1 Vision Sensor

Vision sensors could be further grouped to two types: representations based on local features (Gasparrini et al., 2014; Elangovan et al., 2012) and body skeleton (Wang, Liu, et al., 2012; Shahroudy, Liu, et al., 2016). HAR methods based on local features are independent to the choice of sensors as they only use raw depth data and more robust to occlusion as depth sensors are usually installed on ceiling. In (Gasparrini et al., 2014), approaches using local features could be capable for recognizing simple activities like fall and hand gestures. Another approach (Elangovan et al.) used local features to recognize three types of interactions: human to human, human to object, and human to vehicle, which is at a rough level and has low generalization ability as the local features are fixed. Although local feature-based methods could not provide applicable fine-grained HAR solutions, the performance of a DL method (Haque et al., 2017) proposed to monitor hand hygiene compliance in a hospital outperforms human accuracy.

Comparing with local feature-based approaches, with a skeleton retrieval step, skeleton representation of human activity data could significantly alleviate the complexity of HAR. Vision sensors that allow 2D or 3D skeleton retrieval including:

- Motion Captures (MoCap): MoCaps can provide very accurate skeletal data, but suffered from high price and low flexibility for general usage purpose;
- Depth cameras: some off-the-shelf commercial depth cameras like Microsoft Kinect v1/v2 and Intel RealSense can retrieve skeleton data in a significantly affordable way with acceptable accuracy for HAR;
- RGB cameras: 2D skeleton data could be retrieved from RGB video data (Cao et al., 2016) and even 3D skeleton data (Moreno-Noguer, 2017; Pavlakos et al., 2017), which requires more computational cost and challenging for real-time detection. Detecting 2D skeleton from RGB image resembles the COCO Keypoint Challenge (Lin et al., 2014).

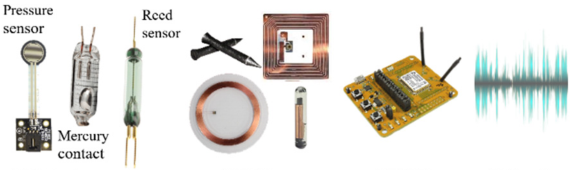
With the capability of collecting skeleton data, Kinect sensors dominate the area of vision-based HAR. According to the list of public benchmark datasets in (Han, Reily, Hoff, & Zhang, 2017), there were 29 out of 41 datasets being collected by Kinect sensors. Mocap ranked the second popular approach in (Han et al., 2017). Other depth sensors like Xtion Live Pro and Leap Motion have seldom been used for data collection (Marin et al., 2014). Representative big HAR datasets collected by Kinect sensors are listed in Table 1. Currently, NTU RGB+D 120 (Liu, Shahroudy, Perez, et al., 2019) has the largest number of activities (NA) and subjects (NS) involved, which is grouped into three categories: daily actions, medicals actions and mutual actions.

Table 1: Benchmark datasets for vision-based HAR.

Dataset	Sensor	NS	NA
(Wang et al., 2012)	Kinect v1	10	16
(Wei et al., 2013)	Kinect v1	8	8
(Rahmani et al., 2014)	Kinect v1	10	30
(Shahroudy et al., 2016)	Kinect v2	40	60
(Liu, Hu, et al., 2017)	Kinect v2	66	51
(Liu, Shahroudy, Perez, et al., 2019)	Kinect v2	106	120

2.2 Ambient Sensor

From the best of our knowledge, there are mainly four types of ambient sensors as shown in Figure 1. Wi-Fi and RFID tags have been claimed to be used for both coarse-grained and fine-grained activity recognition in (Wang, Zhang, et al., 2015) and (Patterson et al., 2005), respectively. While state change sensors are usually only capable for coarse-grained activity recognition. There are also few attempts using audio data for HAR, which is an area separated from speech recognition.



(a) State change sensors (b) RFID tags (c) Wi-Fi (d) Sound

Figure 1: An IMU sensor with 6 degrees of freedom.

2.2.1 State Change Sensors

Various types of state change sensors could be used for collecting ambient state changes of objects like furniture, bathroom sanitary ware, kitchen utensils, and electrical appliance. Figure 1(a) gives three example of commonly used state sensors namely pressure sensor, mercury contact and reed sensor. These state sensors are binary and wireless and can indicate binary states like switch on or off, and open or closed through a Wireless Sensor Network (WSN). For data driven HAR model development and evaluation purposes, self-reporting and camera monitoring are two main methods of data annotation. The HAR method proposed in (Van Kasteren, 2011) used a wireless sensor network kits called RFM DM 1810 to connect various state change sensors includes reed switch, pressure mat, and mercury contact sensor, etc. Except RFM DM 1810, there are other open-source hardware platforms, Udoo and Raspberry Pi which are popular among researchers for their low-cost and highly scalable in terms of both the type and number of sensors. It is concluded in (Maksimović et al., 2014) that the expensive Udoo could achieve the best performance among other IoT hardware platforms including Arduino Uno, BeagleBone Black, Phidgets and Raspberry Pi.

2.2.2 RFID

RFID is a technology for reading information from a distance from RFID-tags. RFID technology can be subdivided into three categories: passive, semi-passive, and active. Depending upon the application, near-field RFID tags come in many form factors as shown in Figure 1(b) (Chawla & Ha, 2007). RFID technology is combined with data-mining techniques and an inferencing engine to recognize activities based on the objects used by people. (Patterson et al., 2005) introduced a fine-grained HAR approach by tagging 60 activity related objects for a morning household routine. A user needs to take a RFID glove as a reader in their approach. A similar job was conducted earlier by (Philipose et al., 2004) for inferring activities of elderlies and ADL collection. RFID technology needs to attach tags to objects and reader devices to users. It is suitable to tag some movable objects as Patterson and Ma did. Meanwhile, it could also be used together with wearable sensors together (Hong et al., 2010; Stikic et al., 2008). The data stream of this approach is similar with state change sensors. In contrast to state change sensors, it is easier to equip small objects, such as a toothbrush or a dinner plate, with a sensing node. State sensors in WSN can therefore only sense activities at relatively limited granularities.

2.2.3 Wi-Fi

Since human bodies are good reflectors of wireless signals, human activities can be recognized by monitoring changes in Wi-Fi signals (Ma et al., 2016). Recent Wi-Fi-based HAR uses the Channel State Information (CSI) of commercial Wi-Fi systems (Guo, 2017). To extract CSI, Intel Wi-Fi Wireless Link 5300 is a frequently used Wi-Fi Network Interface Card (NIC) which supports IEEE 802.11n enabled by modulation methods of OFDM (orthogonal frequency division multiplexing) and MIMO (multiple input multiple output) (Halperin et al., 2011). The movement of the human body parts cause variations in the Wi-Fi signal reflections, which results in changes in CSIs. By analyzing the data streams of CSIs of different activities and comparing them against stored models, human behavior can be recognized. This is done by extracting features from CSI data streams and using machine learning techniques to build models and classifiers. One challenge of this approach is how to make a system robust to environment change. Common device set up is illustrated in Figure 2, which studies the impact of environment differences in (Guo, 2017). Another challenge is multi-user activity recognition, which remains an open question and few solutions have been attempted (Wang, Liu, et al., 2015). With the advantages of low deployment cost, non-intrusive sensing nature, wide coverage range (approximate 70m indoor and 250m outdoor), Wi-Fi based activity recognition has become an emerging and promising research area with the abilities of traverse through wall for HAR and localization of static human subjects and metallic objects (Adib & Katabi, 2013; Pu et al., 2013; Huang, et al., 2014).

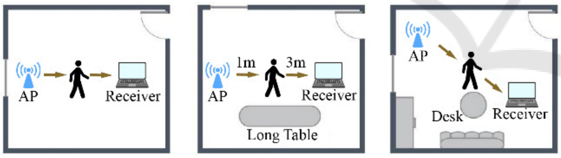


Figure 2: Wi-Fi and PC setups in experimental scenarios.

2.2.4 Sound

In parallel to other ambient sensors, sound produced by objects, human, or human-object interactions convey rich cognitive information about the ongoing context, events, and conversations. Stork et al. (Stork et al., 2012) attempted to recognize activities from audio data by using the Mel Frequency Cepstral Coefficient (MFCC) feature to build a Soundbook for all activities. Stork created the Freiburg dataset as in Table 2. Audio-based HAR is a rarely attempted area separated from speech recognition, hence, there are very few public datasets available. Another relevant dataset is ESC-US Dataset (Piczak, 2015) which has some labelled subsets like ESC-10 and ESC-50.

Representative public datasets for ambient sensor based HAR are listed in Table 2. RFID sensors could be used for fall action only. State-change sensors could be used for some coarse-grained activities, while Wi-Fi and audio signals have the ability for inferring relatively more fine-grained activities.

Table 2: Datasets of ambient sensor-based HAR.

Dataset	Sensor	NS	NA
(Van et al., 2008)	State change	1	8
(Torres et al., 2013; Wickramasinghe, Ranasinghe, et al., 2017)	RFID	14	2
(Wickramasinghe, Torres, et al., 2017)	RFID	13	2
(Guo et al., 2018)	Wi-Fi	10	16
(Stork et al., 2012)	Audio	NA	22

2.3 Wearable Sensor

Wearable inertial sensors like accelerometer and gyroscope have achieved popularity with their advantage of long-term monitoring system (Li, Xu, et al., 2016; Liu, Yen, et al., 2017). (Mukhopadhyay, 2015) introduced the sensors for human activity monitoring could be body temperature, heart rate, accelerometer, and Electrocardiogram (ECG). Except inertial sensors, other physiological condition sensors are more relates to various diseases directly instead of HAR. For example, body temperature and heart rate sensors could be used for detecting health problems like stroke, heart attack and shock. (Parkka et al., 2006) concluded that the accelerometer is the most effective and accurate sensor for HAR. Besides, comparing with accelerometer, the measurement of those physiological condition data is not very relevant to the task of HAR as proved by (Parkka et al., 2006).

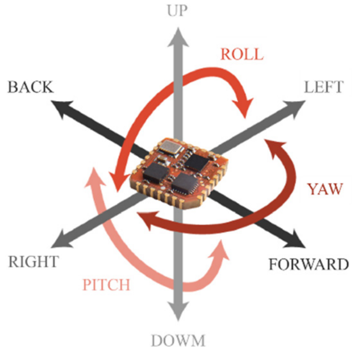


Figure 3: An IMU sensor with 6 degrees of freedom.

An IMU is a Micro-Electro-Mechanical System (MEMS) electronics module and is typically comprised of 3 accelerometers, 3 gyroscopes, and optionally 3 magnetometers. An accelerometer measures changes in velocity and changes in position. A gyroscope measures either changes in orientation or changes in angular velocity. Magnetometers is useful to determine absolute orientation of the sensor. As Figure 3 shows, IMUs with 3 axis accelerometers and 3 axis gyroscopes are commonly referred to as 6 degrees of freedom (DOF) IMU sensors. The inclusion of a 3-axis magnetometer is sometimes referred to as 9 DOF IMU sensors although technically magnetometer should not be referred to as inertial sensor. Table 3 provides some wearable sensor based HAR datasets that use varied IMU sensors from using 3-DOF IMUs to 9-DOF IMUs. All of them are for coarse-grained HAR due to the intrinsic characteristic of IMUs as it could not provide sufficient information like appearance features in the vision sensors.

Table 3: Datasets of wearable sensor-based HAR.

Dataset	Sensor	NS	NA
(Reiss & Stricker, 2012)	3 3-DOF IMUs	9	18
(M. Zhang & Sawchuk, 2012)	6-DOF IMU	14	12
(Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2013)	3-DOF IMU	30	12
(Baños et al., 2012)	9-DOF IMUs	17	33

2.4 Multiple Sensors

Although skeleton-based methods achieved outstanding performances in HAR, common data modalities like color, depth, face, and sound from real-world scenarios have seldom been collectively considered in existing HAR methods. (Chahuara et al., 2016) attempted to fuse sound data with other ambient sensor to recognized human activities in smart homes. For some activities like talking, use mobile phone, typing and eating, audio data might provide some information which is independent with other modalities yet informative for activity recognition. As far as we know, this data modality also has never been attempted together with vision sensor. Given the complementary properties of these signals, multimodal HAR attracts increasing research attention in recent years. Intuitively, multimodal HAR on one hand is more complex to process, on the other hand it contributes the HAR accuracy as diverse sensors can mutually compensate the shortcomings of each other. Table 4 shows some representative multimodal HAR datasets. According to the datasets in Table 4, multimodal methods usually could lead to better activity recognition performance. Hence, to pursue more accurate and higher resolution HAR, multimodal approaches that make use of the complementary advantage of multiple data modalities is the direction of future HAR.

Table 4: Datasets of multimodal HAR.

Dataset	Sensor	NS	NA
(Hodgins & Macey, 2009)	Video, audio, Mocap, 9 IMUs, RFIDs	18	5
(Sagha et al., 2011)	IMUs, 72 sensors of 10 modalities	12	21
(Ofli et al., 2013)	Mocap, Kinect v1, camera, acc, audio	12	11
(Chen, Jafari, et al., 2015)	Kinect v1, 6-DOF IMUs	8	27

3 HAR CAPABILITY OF SENSOR

In this section, we analyze the HAR capability of various IoT sensors by proposing a definition scheme of activity complexity.

3.1 Human Activity Characteristics

From the activity perspective, a clear definition of a human activity complexity is crucial for evaluating the HAR capabilities of different sensors. Human activities vary in terms of many structural characteristics like hierarchical structure, activity duration, location, and the involved number of people or objects. Previous research defined activity complexity by considering the time span only (Bruce & Chan, 2017), which is not adequate to reflect the levels of activity complexity. Low-level activities such as tracking and body posture analysis has been surveyed by (Aggarwal & Cai, 1999). Considering three aspects: object, time, and location, we come up with a human activity categorization scheme as shown in Figure 4.

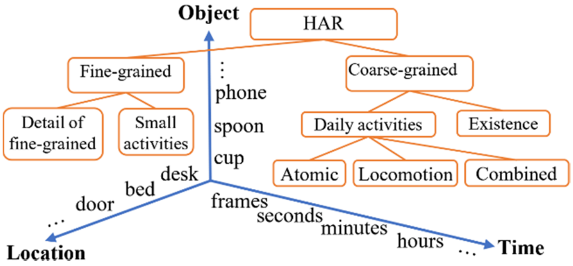


Figure 4: Our human activity categorization scheme that groups activity levels into a hierarchical structure by considering three main activity characteristics include human-object interactions, durations, and locations.

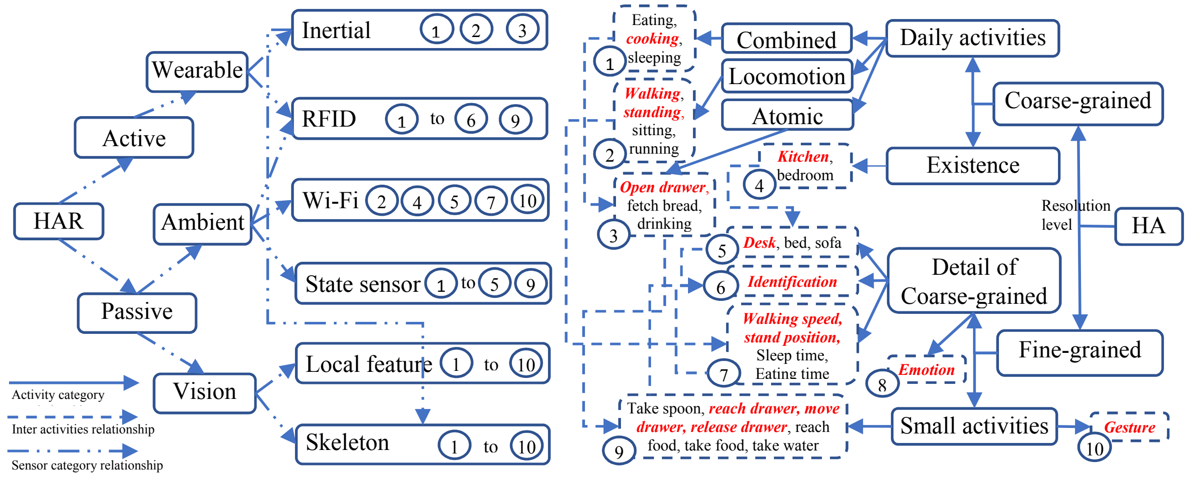


Figure 5: The HAR capability of different IoT sensor-based methods with corresponding to the hierarchical activity categories labeled from 1 to 10. The activity category labels are spread out to the different IoT sensors on the left part of the figure.

3.2 Sensor Capability

Based on the reviewed popular datasets from each IoT sensor category, the HAR capability of different sensors are summarized in Figure 5 with the vision sensor capable for all activity categories (from categories 1 to 10). The activities highlighted with red color and italic font on the right side of Figure 5 are taking the activity examples of a breakfast preparation routine in the job of (Lukowicz et al., 2010). Ambient sensors like the RFID technology used by (Torres et al., 2013; Wickramasinghe, Ranasinghe, et al., 2017; Wickramasinghe, Torres, et al., 2017) needs to install RFID tags on the entire floor of a user’s living environment to detect if the user is near the bed or not. RFID is also affected by noise signals if two objects are very closely located. While the use of state change sensors also needs to install a quite number of sensors to all the related locations and objects, but it could only do some coarse-grained HAR. Wi-Fi CSI is emerged as a novel approach which has the advantage of cross wall sensing ability, but it remains lack of theoretical foundation that proves the measurement accuracy for developing reliable HAR method.

Comparing with ambient sensors, wearable devices could be an appropriate choice for outdoor activity recognition. Given that each sensor modality has its own limitations, it has been surveyed that fusing vision and inertial data improves the accuracy of HAR (Chen, Jafari, et al., 2017). However, the inertial sensor modality does not provide any context information for fine-grained HAR that involves human-object interactions. Besides, due to the intrinsic battery limitation, it is intrusive for users to wear sensor devices do long-term monitoring. One recent trend for multimodal HAR is fusing inertial sensor with vision sensor as reviewed in Section 2.4. However, according to various modality combination results of experiments on Berkeley MHAD, the improvement of the performance by adding more data modalities is very limited (from around 98% to 100%) (Ofli et al., 2013). Sometimes, adding extra modality will even lower the HAR accuracy, which renders the extra modality in vain. The increased problem complexity and affected usability also make multimodal HAR hard to be popularized among end users as well as other stakeholders.

According to the surveyed datasets in Section 2. it is noticeable that NTU RGB+D 120 (Jun Liu, Shahroudy, Perez, et al., 2019) is by far the largest one from perspectives like subjects involved, number of activity classes, and number of viewpoints. Many succeeding jobs have been emerged based on NTU RGB+D 120. Some model activities with a spatial and temporal networks by using CNN and LSTM algorithms (Liu, Shahroudy, Xu, et al., 2016; Song et al., 2017). While some attempt to model the most informative joints in the skeleton data using the context-aware LSTM algorithm (Liu, Wang, et al., 2017) or remove the noise of the skeleton data for view invariant recognition (Zhang et al., 2017; Liu, Liu, et al., 2017). Another potential method is using the contextual information to improve the HAR accuracy by modeling human-object interaction (Wei et al., 2017), which has not been attempted on large datasets like NTU RGB-D 120 or PKU-MMD (Liu, C., et al., 2017).

Given the above comparison, we believe vision sensor could achieve the task of nonintrusive fine-grained HAR. The most controversial concern of privacy could be avoided by using technologies like blurring. Off-the-shelf sensors like RealSense and Kinect could work even in poor illumination condition, which makes them capable for 24-hour activity monitoring. However, for vision sensors being applied to healthcare, it remains some gaps to be conquered as following:

- As surveyed in Section 2, existing algorithm-oriented and performance-oriented jobs usually verify their models' accuracy on benchmark datasets with the activity duration limited to seconds. This means that existing methods has not been applied to applications domains.
- For high resolution HAR, multimodal sensor fusion methods need to be developed to recognize more fine-grained activities that reflect more details of the human behavior.
- For common application domains like healthcare and surveillance, it requires domain experts to validate the feasibility and reliability of the HAR resolution.

4 HAR ALGORITHMS

In the last decade, HAR methods based on single sensor modality have experienced great progress from version-based HAR (Poppe, 2010) to skeleton-based methods (Presti & La Cascia, 2016), and from ambient sensors (Rashidi & Mihailidis, 2013) to wearable sensors (Lara & Labrador, 2013). Since the data stream of sensor-based HAR has sequential features, traditional algorithms like Dynamic Time Warping (DTW) (Gavrila & Davis, 1995), Hidden Markov Model (HMM) (Oliver, Horvitz, & Garg, 2002), and Support Vector Machine (SVM) (Lublinerman et al., 2006) have been commonly used (Aggarwal & Ryoo, 2011), which is recently dominated by DL algorithms (Wang, Chen, et al., 2017). Algorithms for multimodal HAR share a similar trend of using DL models to extract latent features. When it comes to multimodal HAR, sensor fusion is the key issue that needs to be tackled. According to (Elmenreich, 2002), there are three main levels of sensor fusion approaches namely: 1) data-level fusion, 2) feature-level fusion, and 3) decision-level fusion, which is illustrated in Figure 6. Multimodal methods have been attempted by both traditional methods with a feature extraction step and DL methods with end-to-end training manners. In this section, we introduce both the traditional and DL models form perspectives of single modal and multimodal methods.

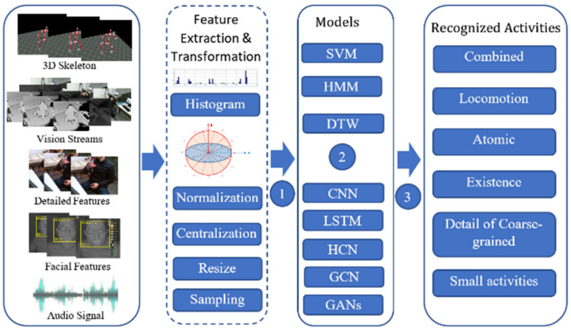


Figure 6: Common workflow of sensor-based HAR. Three sensor fusion methods data-level fusion, feature-level fusion, and decision-level fusion are labeled as 1, 2 and 3, respectively.

4.1 Traditional HAR Algorithms

Since input data types are intrinsically heterogeneous, sensor fusion at data-level has seldom been attempted by researchers. Sensor fusion conducted at feature-level calculates popular features from input data and further combines them into a fused feature vector for inferring activity classes. For example, (Liu, Yang, et al., 2010) fused quantized vocabulary of local spatial-temporal volumes (cuboids and 2-D SIFT) and the higher-order statistical models of interest points for activity recognition using a hyper-sphere multi-class SVM. Decision-level fusion uses multiple classifiers for corresponding multiple features and makes the classification by considering the complementary results among classifiers. For instance, (Tran et al., 2010) proposed a baseline approach using disparate spatial features as an input vector to train multiple HMM models within a fusion framework. Similar fusion approach also used in skeleton-based method by (Xia et al., 2012) that leans an HMM model for each activity.

Traditional machine learning algorithms like SVM, kernel machines, discriminant analysis, and spectral clustering, concatenate all multiple views into a single view to fit their learning settings. However, this concatenation is not physically meaningful as each view has specific statistical properties and usually causes overfitting in case of small dataset size. In contrast, multi-view learning as another paradigm which uses one function to model a particular view and simultaneously optimizes all the functions to exploit the redundant views of the same input data and improve the learning performance. According to the categorization in (Xu et al., 2013), multi-view learning is categorized into three main types namely co-training, Multiple Kernel Learning (MKL), and subspace learning.

Co-training was originally proposed for the problem of semi-supervised learning, in which there is access to labelled as well as unlabelled data. It considers a setting in which each example can be partitioned into two distinct views and makes three main assumptions for its success: sufficiency, compatibility, and conditional independence. The original co-training job described experiments using co-training to classify web pages into "academic course home page" or not (Blum & Mitchell, 1998). The classifier correctly categorized 95% of 788 web pages with only 12 labelled web pages as training data. Co-training is famous for its capability for automatically learning two independent and sufficient representations from data when only small amounts of labelled data and large amounts of unlabelled data are available. From the best of our knowledge, seldom existing HAR tasks using the semi-supervised co-training method. MKL has been used by (Althloothi et al., 2014) and (Ofli et al., 2013). MKL method fuses at kernel level to select useful features based on weights. (Althloothi et al., 2014) uses motion features and shape features of skeleton and depth image to train multiclass-SVM based classifiers for activity recognition. Another popular MKL algorithm is Adaptive Boosting (AdaBoost) (Lv & Nevatia, 2006), which relies on constructing effective geometric features for improving the HAR accuracy.

Subspace learning-based approaches aim to obtain a latent subspace shared by multiple views by assuming that the input views are generated from this subspace. The well know subspace learning-based approach is Canonical Correlation Analysis (CCA) (Hardoon et al., 2004) and KCCA (Lai & Fyfe, 2000), which gives the correlated form of input modalities as a robust representation of multimodal data through linear projections. The purpose of Correlation-Independence Analysis (CIA) is to identify the strength of respective modalities through teasing out their common and independent components and to utilize them for improving the classification accuracy of human activities.

4.2 Deep Learning Approaches

4.2.1 Single Modal Approach

Recently, the advance of DL makes it possible to perform automatic high-level feature extraction thus achieves promising performance in many areas. Since then, DL based methods have been widely adopted for various sensor-based HAR tasks. (Wang et al., 2017) reviewed DL models for HAR tasks, which includes Deep neural network (DNN), Convolutional Neural Network (ConvNets, or CNN), Stacked autoencoder (SAE) etc. Representative DL models are listed in Table 5. (Hammerla et al., 2016) investigated the performance of DNN, CNN and RNN through 4,000 experiments on some public HAR datasets with a conclusion that RNN and LSTM are recommended to recognize short activities that have natural order while CNN is better at inferring long term repetitive activities. The reason is that RNN could make use of the time-order relationship between sensor readings, and CNN is more capable of learning deep features contained in recursive patterns. (Zheng et al., 2014) summarized that CNN is better for multimodal signals.

Table 5: Deep Learning models for HAR tasks.

DL Model	Description
DNN	Deep fully connected network, artificial neural network with deep layers.
CNN	Convolutional neural network, multiple convolution operations for feature extraction.
RNN	Recurrent neural network, network with time correlations and LSTM.
DBN/RBM	Deep belief network and restricted Boltzmann machine.
SAE	Stacked autoencoder, feature learning by decoding-encoding autoencoder.
HCN	Hierarchical cooccurrence network.
GCN	Graph Convolutional Network.
Hybrid	Combination of some deep models.

Concerning the popularity and capability of vision sensor, we further examine the algorithms of vision-based approaches. Existing researches of vision-based HAR methods mainly focus on three directions for the improvement of single modality HAR. The first direction focuses on data pre-processing and data cleaning. For example, (Liu, M., et al., 2017) proposed a method that remove the noise data in skeleton activity representations by learning a model that reconstructs more accurate skeleton data. An approach with this motivation has also been proposed by (Zhang et al., 2017).

The second approach improves the HAR benchmarks by proposing novel learning or representing models. (Liu, Wang, et al., 2017) proposed a context aware LSTM model that could learn which part of joints contribute the HAR. (Yan et al., 2018) introduced a Spatial Temporal-Graph Convolutional Network (ST-GCN) that learns both the spatial and temporal patterns from skeleton-based activity data. Many enhanced versions of GCN models has been proposed that improve the ST-GCN by considering other physical prior knowledge. For example, (Shi et al., 2018) proposed a non-local GCN that leans the graph structure individually for different layers and samples and achieved improved performance than the manually designed convolutional operation of ST-GCN. Another GCN method proposed by (Li et al., 2019) tries to model discriminative features from actional and structural links of the skeleton graph. Except GCN, motivated by cooccurrence learning, (Li et al., 2018) proposed the hierarchical cooccurrence network (HCN) that learns point-level features aggregated to cooccurrence features with a hierarchical methodology. The co-occurrence features refer to the interactions and combinations of some subsets of skeleton joints that characterizes an action (Zhu et al., 2016). Considering both the graph and cooccurrence characteristics, (Si et al., 2019) proposed an Attention Enhanced Graph Convolutional LSTM Network (AGC-LSTM) that achieved high accuracy on the NTU-RGB-D dataset. However, the Directed Graph Network (DGN) (Shi et al., 2019a) and achieved higher accuracy than AGC-LSTM on the NTU-RGB-D dataset with a small margin.

The third method is data augmentation that leans data generation models to produce more training data for the purpose of feeding more fuel to deep learning models. (Barsoum et al., 2017) developed a sequence-to-sequence model for probabilistic human motion prediction, which predicts multiple plausible future human poses from the same input. However, it has not yet been evaluated if the generated data could be used for improving the generalization power or accuracy of HAR tasks. Focusing on the improvement of accuracy on benchmark datasets might neglect the improvement of HAR itself and its application domains that needs higher activity resolution. Another neglected issue of existing methods is segmentation as deploying HAR methods to domain applications needs simultaneously performing both tasks of segmentation and classification.

4.2.2 Multimodal Approach

From the confusion matrix comparison of (Si et al., 2019), skeleton modality cannot provide sufficient information to discriminate action pairs that include human-object interactions like “reading” and “writing”, “writing” and “typing on a keyboard”, “pointing to something with finger” and “pat on back of other person”, which is due to the similar skeleton motion patterns shared by those activity pairs. Similarly, for activities that include interaction with items in the PKU-MMD, skeleton modality might not be capable to provide sufficient features from the interacted items. Intuitively, these difficult activity pairs have higher resolution than the well classified ones, which might need more detailed contextual or semantic features from other data modalities like RGB and depth channels of the datasets like PKU-MMD and NTU RGB-D.

The multimodal fusion analysis of (Ordóñez & Roggen, 2016) on the Opportunity dataset (Sagha et al., 2011) indicates that feeding more data channels to its model called DeepConvLSTM would get performance improvement. Similarly, experimental results of (Jun Liu, Shahroudy, Perez, et al., 2019) on NTU RGB+D 120 also indicates that extra data modalities contribute the classification accuracy. Hence, it is commonly accepted that multimodal HAR approaches have the potential to improve the activity resolution and recognize difficult activities. Existing multimodal HAR methods could be roughly categorized to two classes: vision-based multimodal (Wei et al., 2017; Shahroudy, Ng, et al., 2017; Wu et al., 2016) and vision-wearable based multimodal (Sagha et al., 2011; Chen, Jafari, et al., 2015); (Ordóñez & Roggen, 2016). The 4DHOI model proposed by (Wei et al., 2017) attempts to represent both 3D human poses and contextual objects in events by using a hierarchical spatial temporal graph. The fusion concept of (Wu et al.) has two approaches namely late fusion and intermediate fusion. The late approach simply combines the emission probabilities from two modalities. In the intermediate fusion scheme, each modality (skeleton and RGB-D) is first pretrained separately, and their high-level representation are concatenated to generate a shared representation. The HAR tasks in (Wei et al., 2017) include segmentation, recognition, and object localization, which is not just recognition as what is doing by most of the latest skeleton-based solutions (Li, Chen, et al., 2019; Si et al., 2019; Shi et al., 2019a; Liang et al., 2019) and (Shi et al., 2019b). With segmentation, the solution might be capable of doing HAR in an online mode. Another online skeleton-based HAR solution was proposed by (Liu, Shahroudy, Wang, et al., 2019).

5 CONCLUSIONS

This survey investigated IoT sensors for HAR that could be applied to application domains like habit perception, intervention performance evaluation, disease prediction, and adaptive (automatic) smart home. It provides a systematic view of the HAR domain by disentangling the IoT sensors utilized and their corresponding status in terms of their datasets and algorithms. By reviewing the datasets of each sensor modality and proposing a human activity categorization scheme that groups human activities based on their levels of complexity, we analyzed the HAR capability of different IoT sensors and concluded that vision sensors are relatively more capable for HAR tasks. For HAR algorithms, we investigated both traditional algorithms and DL models. It is worth to note that the hard cases in the NTU-RGB+D dataset could not be well recognized by the skeleton modality only. Multimodal method has the potential of recognizing more fine-grained activities, but existing algorithms remain uncapable to tackle multimodal data well.

For future jobs, as we summarized research gaps in Section 3, although increasingly larger datasets were collected, higher resolution HAR methods need to be developed for landing domain applications like healthcare, assistive technologies, and surveillance. To do so, multimodal methods being capable of higher resolution HAR need to be developed in the future. Besides, the involvement of domain experts is essential to validate the feasibility and reliability of future HAR methods and applications.

REFERENCES

Acampora, G., Cook, D. J., Rashidi, P., & Vasilakos, A. V. (2013). A survey on ambient intelligence in healthcare. *Proceedings of the IEEE, 101*(12), 2470-2494.

Adib, F., & Katabi, D. (2013). *See through walls with WiFi!* (Vol. 43): ACM.

Aggarwal, J. K., & Cai, Q. (1999). Human motion analysis: A review. *Computer Vision and Image Understanding, 73*(3), 428-440.

Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR), 43*(3), 16.

Alemdar, H., & Ersoy, C. (2010). Wireless sensor networks for healthcare: A survey. *Computer networks, 54*(15), 2688-2710.

Althloothi, S., Mahoor, M. H., Zhang, X., & Voyles, R. M. (2014). Human activity recognition using multi-features and multiple kernel learning. *Pattern Recognition, 47*(5), 1800-1812.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013). *A Public Domain Dataset for Human Activity Recognition using Smartphones.* Paper presented at the ESANN.

Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., & Havinga, P. (2010). *Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey.* Paper presented at the Architecture of computing systems (ARCS), 2010 23rd international conference on.

Baños, O., Damas, M., Pomares, H., Rojas, I., Tóth, M. A., & Amft, O. (2012). *A benchmark dataset to evaluate sensor displacement in activity recognition.* Paper presented at the Proceedings of the 2012 ACM Conference on Ubiquitous Computing.

Barsoum, E., Kender, J., & Liu, Z. (2017). HP-GAN: Probabilistic 3D human motion prediction via GAN. *arXiv preprint arXiv:1711.09561.*

Blum, A., & Mitchell, T. (1998). *Combining labeled and unlabeled data with co-training.* Paper presented at the Proceedings of the eleventh annual conference on Computational learning theory.

Bruce, X., & Chan, K. C. (2017). *Discovering Knowledge by Behavioral Analytics for Elderly Care.* Paper presented at the Big Knowledge (ICBK), 2017 IEEE International Conference on.

Bulling, A., Blanke, U., & Schiele, B. (2014). A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR), 46*(3), 33.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2016). Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *arXiv preprint arXiv:1611.08050.*

Chahuara, P., Fleury, A., Portet, F., & Vacher, M. (2016). On-line human activity recognition from audio and home automation sensors: Comparison of sequential and non-sequential models in realistic Smart Homes 1. *Journal of ambient intelligence and smart environments, 8*(4), 399-422.

Chawla, V., & Ha, D. S. (2007). An overview of passive RFID. *IEEE Communications Magazine, 45*(9).

Chen, C., Jafari, R., & Kehtarnavaz, N. (2015). *Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor.* Paper presented at the Image Processing (ICIP), 2015 IEEE International Conference on.

Chen, C., Jafari, R., & Kehtarnavaz, N. (2017). A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications, 76*(3), 4405-4425.

Chen, L., Hoey, J., Nugent, C. D., Cook, D. J., & Yu, Z. (2012). Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42*(6), 790-808.

Chouhan, V. L., & Sharma, P. (2017). Behavioral Interventions in Autism.

Dilip Kumar, S., WA (US); Elie Micah Kornield, Seattle, WA (US); Alexander Clark Prater, Seattle, WA (US); Sridhar Boyapati, Sammamish, WA (U S); Xiaofeng Ren, Sammamish, WA (US); Chang Yuan, Seattle, WA (Us). (2015). United States Patent No. US 20150019391A1. Patent Application Publication.

Elangovan, V., Bandaru, V. K., & Shirkhodaie, A (2012). Team activity analysis and recognition based on Kinect depth map and optical imagery techniques, Proceedings Volume 8392, Signal Processing, Sensor Fusion, and Target Recognition XXI; 83920W

Elmenreich, W. (2002). An Introduction to Sensor Fusion. *Vienna University of Technology, Austria.*

Gasparrini, S., Cippitelli, E., Spinsante, S., & Gambi, E. (2014). A depth-based fall detection system using a Kinect® sensor. *Sensors, 14*(2), 2756-2775.

Gavrila, D. M., & Davis, L. S. (1995). *Towards 3-d model-based tracking and recognition of human movement: a multi-view approach.* Paper presented at the International workshop on automatic face-and gesture-recognition.

Gianna Lise Puerini, B., WA (US); Dilip Kumar, Seattle, WA (US); Steven Kessel, Seattle, WA (US). (2015). United States Patent No. US 20150012396A1. Patent Application Publication.

Guo, L., Wang, L., Liu, J., Zhou, W., & Lu, B. (2018). HuAc: Human Activity Recognition Using Crowdsourced WiFi Signals and Skeleton Data. *Wireless Communications and Mobile Computing, 2018*.

Halperin, D., Hu, W., Sheth, A., & Wetherall, D. (2011). Tool release: Gathering 802.11 n traces with channel state information. *ACM SIGCOMM Computer Communication Review, 41*(1), 53-53.

Hammerla, N. Y., Halloran, S., & Ploetz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880*.

Han, F., Reily, B., Hoff, W., & Zhang, H. (2017). Space-time representation of people based on 3D skeletal data: A review. *Computer Vision and Image Understanding, 158*, 85-105.

Haque, A., Guo, M., Alahi, A., Yeung, S., Luo, Z., Rege, A., . . . Singh, A. (2017). Towards Vision-Based Smart Hospitals: A System for Tracking and Monitoring Hand Hygiene Compliance. *arXiv preprint arXiv:1708.00163*.

Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation, 16*(12), 2639-2664.

Hodgins, F., & Macey, J. (2009). Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. *CMU-RI-TR-08-22*.

Hong, Y.-J., Kim, I.-J., Ahn, S. C., & Kim, H.-G. (2010). Mobile health monitoring system based on activity recognition using accelerometer. *Simulation Modelling Practice and Theory, 18*(4), 446-455.

Huang, D., Nandakumar, R., & Gollakota, S. (2014). *Feasibility and limits of wi-fi imaging.* Paper presented at the Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems.

Lai, P. L., & Fyfe, C. (2000). Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems, 10*(05), 365-377.

Lara, O. D., & Labrador, M. A. (2013). A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials, 15*(3), 1192-1209.

Li, C., Zhong, Q., Xie, D., & Pu, S. (2018). Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation. *arXiv preprint arXiv:1804.06055*.

Li, M., Chen, S., Chen, X., Zhang, Y., Wang, Y., & Tian, Q. (2019). *Actional-Structural Graph Convolutional Networks for Skeleton-based Action Recognition.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Li, X., Xu, H., & Cheung, J. T. (2016). Gait-force model and inertial measurement unit-based measurements: A new approach for gait analysis and balance monitoring. *Journal of Exercise Science & Fitness, 14*(2), 60-66.

Liang, D., Fan, G., Lin, G., Chen, W., Pan, X., & Zhu, H. (2019). *Three-Stream Convolutional Neural Network With Multi-Task and Ensemble Learning for 3D Action Recognition.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014). *Microsoft coco: Common objects in context.* Paper presented at the European conference on computer vision.

Linlin Guo, L. W., Jialin Liu, Wei Zhou, Bingxian Lu, Tao Liu, Guangxu Li, Chen Li. (2017). *A novel benchmark on human activity recognition using WiFi signals*. Paper presented at the 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), Dalian, China.

Liu, C., Hu, Y., Li, Y., Song, S., & Liu, J. (2017). *PKU-MMD: A Large Scale Benchmark for Skeleton-Based Human Action Understanding.* Paper presented at the Proceedings of the Workshop on Visual Analysis in Smart and Connected Communities.

Liu, J., Shahroudy, A., Perez, M. L., Wang, G., Duan, L.-Y., & Chichung, A. K. (2019). NTU RGB+ D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. *IEEE transactions on pattern analysis and machine intelligence*.

Liu, J., Shahroudy, A., Wang, G., Duan, L.-Y., & Chichung, A. K. (2019). Skeleton-Based Online Action Prediction Using Scale Selection Network. *IEEE transactions on pattern analysis and machine intelligence*.

Liu, J., Shahroudy, A., Xu, D., & Wang, G. (2016). *Spatio-temporal lstm with trust gates for 3d human action recognition.* Paper presented at the European Conference on Computer Vision.

Liu, J., Wang, G., Hu, P., Duan, L.-Y., & Kot, A. C. (2017). *Global context-aware attention lstm networks for 3d action recognition.* Paper presented at the CVPR.

Liu, J., Yang, J., Zhang, Y., & He, X. (2010). *Action recognition by multiple features and hyper-sphere multi-class svm.* Paper presented at the Pattern Recognition (ICPR), 2010 20th International Conference on.

Liu, K.-C., Yen, C.-Y., Chang, L.-H., Hsieh, C.-Y., & Chan, C.-T. (2017). *Wearable sensor-based activity recognition for housekeeping task.* Paper presented at the Wearable and Implantable Body Sensor Networks (BSN), 2017 IEEE 14th International Conference on.

Liu, M., Liu, H., & Chen, C. (2017). Enhanced skeleton visualization for view invariant human action recognition. *Pattern Recognition, 68*, 346-362.

Lublinerman, R., Ozay, N., Zarpalas, D., & Camps, O. (2006). *Activity recognition from silhouettes using linear systems and model (in) validation techniques.*

Paper presented at the 18th International Conference on Pattern Recognition (ICPR'06).

Lukowicz, P., Pirkl, G., Bannach, D., Wagner, F., Calatroni, A., Förster, K., . . . Tröster, G. (2010). *Recording a complex, multi modal activity data set for context recognition.* Paper presented at the Architecture of Computing Systems (ARCS), 2010 23rd International Conference on.

Lv, F., & Nevatia, R. (2006). *Recognition and segmentation of 3-d human action using hmm and multi-class adaboost.* Paper presented at the European conference on computer vision.

Ma, J., Wang, H., Zhang, D., Wang, Y., & Wang, Y. (2016). *A Survey on Wi-Fi Based Contactless Activity Recognition.* Paper presented at the 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/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences.

Maksimović, M., Vujović, V., Davidović, N., Milošević, V., & Perišić, B. (2014). Raspberry Pi as Internet of things hardware: performances and constraints. *design issues, 3*, 8.

Marin, G., Dominio, F., & Zanuttigh, P. (2014). *Hand gesture recognition with leap motion and kinect devices.* Paper presented at the Image Processing (ICIP), 2014 IEEE International Conference on.

Moreno-Noguer, F. (2017). *3d human pose estimation from a single image via distance matrix regression.* Paper presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Mukhopadhyay, S. C. (2015). Wearable sensors for human activity monitoring: A review. *IEEE sensors journal, 15*(3), 1321-1330.

Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., & Bajcsy, R. (2013). *Berkeley MHAD: A comprehensive multimodal human action database.* Paper presented at the Applications of Computer Vision (WACV), 2013 IEEE Workshop on.

Oliver, N., Horvitz, E., & Garg, A. (2002). *Layered representations for human activity recognition.* Paper presented at the Proceedings. Fourth IEEE International Conference on Multimodal Interfaces.

Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors, 16*(1), 115.

Palumbo, F., Ullberg, J., Štimec, A., Furfari, F., Karlsson, L., & Coradeschi, S. (2014). Sensor network infrastructure for a home care monitoring system. *Sensors, 14*(3), 3833-3860.

Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., & Korhonen, I. (2006). Activity classification using realistic data from wearable sensors. *IEEE Transactions on information technology in biomedicine, 10*(1), 119-128.

Patterson, D. J., Fox, D., Kautz, H., & Philipose, M. (2005). *Fine-grained activity recognition by aggregating abstract object usage.* Paper presented at the Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on.

Pavlakos, G., Zhou, X., Derpanis, K. G., & Daniilidis, K. (2017). *Coarse-to-fine volumetric prediction for single-image 3D human pose.* Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on.

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J., & DeKosky, S. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review) Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology, 56*(9), 1133-1142.

Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H., & Hahnel, D. (2004). Inferring activities from interactions with objects. *IEEE Pervasive Computing, 3*(4), 50-57.

Piczak, K. J. (2015). *ESC: Dataset for environmental sound classification.* Paper presented at the Proceedings of the 23rd ACM international conference on Multimedia.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing, 28*(6), 976-990.

Presti, L. L., & La Cascia, M. (2016). 3D skeleton-based human action classification: A survey. *Pattern Recognition, 53*, 130-147.

Pu, Q., Gupta, S., Gollakota, S., & Patel, S. (2013). *Whole-home gesture recognition using wireless signals.* Paper presented at the Proceedings of the 19th annual international conference on Mobile computing & networking.

Rahmani, H., Mahmood, A., Huynh, D. Q., & Mian, A. (2014). *HOPC: Histogram of oriented principal components of 3D pointclouds for action recognition.* Paper presented at the European Conference on Computer Vision.

Rashidi, P., & Mihailidis, A. (2013). A survey on ambient-assisted living tools for older adults. *IEEE journal of biomedical and health informatics, 17*(3), 579-590.

Reiss, A., & Stricker, D. (2012). *Creating and benchmarking a new dataset for physical activity monitoring.* Paper presented at the Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments.

Sagha, H., Digumarti, S. T., Millán, J. d. R., Chavarriaga, R., Calatroni, A., Roggen, D., & Tröster, G. (2011). *Benchmarking classification techniques using the Opportunity human activity dataset.* Paper presented at the Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on.

Shahroudy, A., Liu, J., Ng, T.-T., & Wang, G. (2016). *NTU RGB+ D: A large scale dataset for 3D human activity analysis.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Shahroudy, A., Ng, T.-T., Gong, Y., & Wang, G. (2017). Deep multimodal feature analysis for action recognition in rgb+ d videos. *IEEE transactions on pattern analysis and machine intelligence*.

Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2018). Adaptive Spectral Graph Convolutional Networks for Skeleton-

Based Action Recognition. *arXiv preprint arXiv:1805.07694*.

Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019a). *Skeleton-Based Action Recognition With Directed Graph Neural Networks.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019b). *Two-Stream Adaptive Graph Convolutional Networks for Skeleton-Based Action Recognition.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Si, C., Chen, W., Wang, W., Wang, L., & Tan, T. (2019). An Attention Enhanced Graph Convolutional LSTM Network for Skeleton-Based Action Recognition. *arXiv preprint arXiv:1902.09130*.

Song, S., Lan, C., Xing, J., Zeng, W., & Liu, J. (2017). *An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data.* Paper presented at the AAAI.

Stikic, M., Huynh, T., Van Laerhoven, K., & Schiele, B. (2008). *ADL recognition based on the combination of RFID and accelerometer sensing.* Paper presented at the Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on.

Stork, J. A., Spinello, L., Silva, J., & Arras, K. O. (2012). *Audio-based human activity recognition using non-markovian ensemble voting.* Paper presented at the RO-MAN, 2012 IEEE.

Torres, R. L. S., Ranasinghe, D. C., Shi, Q., & Sample, A. P. (2013). *Sensor enabled wearable RFID technology for mitigating the risk of falls near beds.* Paper presented at the RFID (RFID), 2013 IEEE International Conference on.

Tran, K., Kakadiaris, I. A., & Shah, S. K. (2010). *Fusion of human posture features for continuous action recognition.* Paper presented at the European Conference on Computer Vision.

Van Kasteren, T., Noulas, A., Englebienne, G., & Kröse, B. (2008). *Accurate activity recognition in a home setting.* Paper presented at the Proceedings of the 10th international conference on Ubiquitous computing.

Van Kasteren, T. L. M. (2011). *Activity recognition for health monitoring elderly using temporal probabilistic models*.

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2017). Deep Learning for Sensor-based Activity Recognition: A Survey. *arXiv preprint arXiv:1707.03502*.

Wang, J., Liu, Z., Wu, Y., & Yuan, J. (2012). *Mining actionlet ensemble for action recognition with depth cameras.* Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.

Wang, T., Zhang, D., Wang, Z., Jia, J., Ni, H., & Zhou, X. (2015). *Recognizing gait pattern of Parkinson's disease patients based on fine-grained movement function features.* Paper presented at the Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015 IEEE 12th Intl Conf on.

Wang, W., Liu, A. X., Shahzad, M., Ling, K., & Lu, S. (2015). *Understanding and modeling of wifi signal based human activity recognition.* Paper presented at the Proceedings of the 21st annual international conference on mobile computing and networking.

Wei, P., Zhao, Y., Zheng, N., & Zhu, S.-C. (2013). *Modeling 4d human-object interactions for event and object recognition.* Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

Wei, P., Zhao, Y., Zheng, N., & Zhu, S.-C. (2017). Modeling 4D human-object interactions for joint event segmentation, recognition, and object localization. *IEEE transactions on pattern analysis and machine intelligence, 39*(6), 1165-1179.

Wickramasinghe, A., Ranasinghe, D. C., Fumeaux, C., Hill, K. D., & Visvanathan, R. (2017). Sequence learning with passive RFID sensors for real-time bed-egress recognition in older people. *IEEE journal of biomedical and health informatics, 21*(4), 917-929.

Wickramasinghe, A., Torres, R. L. S., & Ranasinghe, D. C. (2017). Recognition of falls using dense sensing in an ambient assisted living environment. *Pervasive and mobile computing, 34*, 14-24.

Wu, D., Pigou, L., Kindermans, P.-J., Le, N. D.-H., Shao, L., Dambre, J., & Odobez, J.-M. (2016). Deep dynamic neural networks for multimodal gesture segmentation and recognition. *IEEE transactions on pattern analysis and machine intelligence, 38*(8), 1583-1597.

Xia, L., Chen, C.-C., & Aggarwal, J. K. (2012). *View invariant human action recognition using histograms of 3d joints.* Paper presented at the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.

Xu, C., Tao, D., & Xu, C. (2013). A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*.

Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. *arXiv preprint arXiv:1801.07455*.

Zhang, M., & Sawchuk, A. A. (2012). *USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors.* Paper presented at the Proceedings of the 2012 ACM Conference on Ubiquitous Computing.

Zhang, P., Lan, C., Xing, J., Zeng, W., Xue, J., & Zheng, N. (2017). View adaptive recurrent neural networks for high performance human action recognition from skeleton data. *arXiv, no. Mar.*

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). *Time series classification using multi-channels deep convolutional neural networks.* Paper presented at the International Conference on Web-Age Information Management.

Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., & Xie, X. (2016). *Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks.* Paper presented at the 30th AAAI Conference on Artificial Intelligence.