This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Sensing Technology for Human Activity Recognition: a Comprehensive Survey

BIYING FU[1, 2], NASER DAMER[1, 2], (MEMBER, IEEE), FLORIAN KIRCHBUCHNER[1], AND ARJAN KUIJPER.[1, 2], (Member, IEEE)

[1]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, 64283 Germany (e-mail: FirstName.LastName@igd.fraunhofer.de)
[2]Interactive Graphics Systems Group, Technical University of Darmstadt, Darmstadt, 64283 Germany

Corresponding author: Biying Fu (e-mail: biying.fu@igd.fraunhofer.de).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT Sensors are devices that quantify the physical aspects of the world around us. This ability is important to gain knowledge about human activities. Human Activity recognition plays an import role in people's everyday life. In order to solve many human-centered problems, such as health-care, and individual assistance, the need to infer various simple to complex human activities is prominent. Therefore, having a well defined categorization of sensing technology is essential for the systematic design of human activity recognition systems. By extending the sensor categorization proposed by White, we survey the most prominent research works that utilize different sensing technologies for human activity recognition tasks. To the best of our knowledge, there is no thorough sensor-driven survey that considers all sensor categories in the domain of human activity recognition with respect to the sampled physical properties, including a detailed comparison across sensor categories. Thus, our contribution is to close this gap by providing an insight into the state-of-the-art developments. We identify the limitations with respect to the hardware and software characteristics of each sensor category and draw comparisons based on benchmark features retrieved from the research works introduced in this survey. Finally, we conclude with general remarks and provide future research directions for human activity recognition within the presented sensor categorization.

INDEX TERMS sensor categorization, human activity recognition, public databases for human activity recognition, physical sensors, sensor benchmark

I. INTRODUCTION

"In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind" by Lord Kelvin (William Thomson) [1]. Sensors are devices that can help to detect and quantify physical aspects of the world around us. They can measure the intensity of light, translate the degree of heat into temperature, or turn mechanical pressure into a force quantity. Sensors are all around us. One of the highest rates of growth of sensor deployment have been in the automotive sector. A modern automobile is equipped with an average of 60 to 100 sensing devices with a rising trend mainly for functional aspects, such as the engine operation, brakes, safety, or emission controls [2]. With the growing trend of smart vehicles, the demand on more sensing units is expected. Also in the smart home domain, miniaturized sensing devices are widespread. The distributed sensors build up an invisible wireless network connecting everything together.

In order to facilitate a sensor comparison and obtaining a comprehensive overview of the sensing technology, researchers try to categorize them into different categories. Sensor classification scheme can range in its complexity. Simple general schemes commonly conclude three sensor categories based on the nature of the sensed property (physical, chemical, and biological) [3]. However, a more complex categorization is often required when addressing distinguished applications. This work focus on the sensing technology deployed in academic research and consumer products for Human Activity Recognition (HAR). To build our sensor categorization within this field, we adopt the classification scheme proposed by White [3]. This scheme is accredited to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

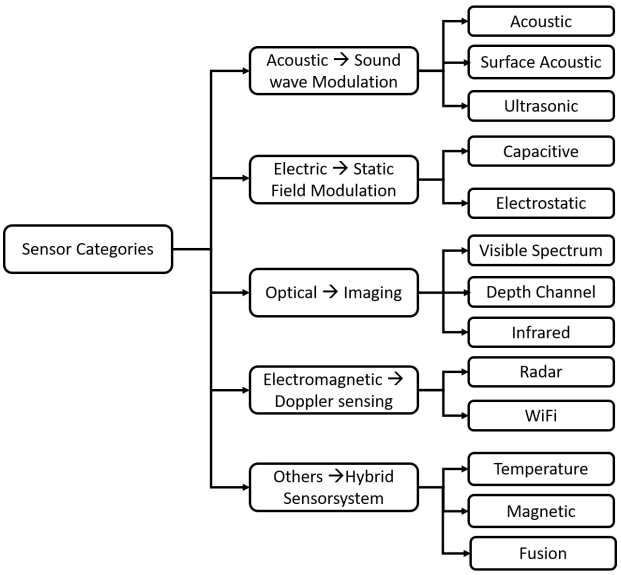

FIGURE 1. The sensor categorization for HAR as presented in this work and based (at the first categorization level) on the work by White [3]. We further extended this definition to include the measuring methods, commonly used in the domain of HAR.

be more flexible and intermediate in complexity. It is based on the measurands or physical entity that a sensor actually senses such as temperature, light intensity, or mechanical stress. We present a first look at our categorization scheme in Figure 1, where we show the first level categorization based on the physical quantities followed by common sensor types utilized to measure this appropriate physical quantity.

Although several surveys have been conducted for HAR for specific sensor categories, such as surveys on acceleration-based [4], [5], radar-based [6], radio-based HAR [7] and camera-based HAR [8], these are all focusing on single sensor technology based applications for a sub-domain of HAR. A thorough comparison across these sensing technique categories with a focus on the sensor advantages and disadvantages in specific tasks is still lacking. Other surveys focus on algorithm-based methods (recent advances made in deep learning [9], [10] and transfer learning [11] applied in the domain of HAR). Hussain [12] combined several surveys and proposed the first survey covering almost all the sub-fields of activity recognition using device-free sensors. However, this work was application-driven (rather than sensor-driven) and largely focused on RFID technology in activity recognition. Unlike other surveys regarding tag-based RFID applications, they promoted the current development of using RFID as device-free solutions for HAR. In contrast to previous works, we are presenting a wide sensor-driven overview on HAR without limitation to a certain sub-application or a certain sensing technology (e.g. ambient sensors). Instead of counting sensor technologies on specific sub-domains of HAR and thus under representing certain sensor categories, we categorize sensors based on its physical properties to adjudge its membership to sub-domains of HAR. Tasks may differ, but the sensor physical characteristics remain. The appropriate sensor category to use is left as a design choice to the application designers. Based on this survey, the application designers should be able to consider the appropriate sensor category with respect to specific task. This survey provides useful insight for researchers and developers in the HAR domain and provides a summary of existing works, including insight into the current and future research directions.

This manuscript is organized as follows. In section II, we present our sensor categorization scheme according to the physical entity they measure and revise the most prominent works utilizing these sensor categories in the domain of HAR. In section III, we provide a detailed discussion of public available databases intend to help developing applications in this research domain with the corresponding sensor categories. In section IV, we present the common evaluation metrics used in the literature to evaluate and compare the performance of the developed algorithms and systems. This is followed by a thorough discussion (Section V) on the hardware and software limitations we identified for each sensor category based on the literature research conducted within this work. Finally, in section VI, we provide the reader with insight into possible solutions to the previously mentioned challenges and offer an overview on current and upcoming future research directions in the domain of HAR with sensory data.

II. SENSORS

A sensor is in general a converter that turns a physical quantity into electric values to be perceived by a digital system. Its output changes according to the change of physical properties on the input side. Sensors integrated in smart environments can either unobtrusively perceive the environment or be directly interacted with. Sensors that tend to sense the natural human intention without direct interaction can be used to design implicit interaction interfaces. Sensors that expect the user to initiate a direct interaction is used to design explicit interaction interface. To choose the appropriate sensor type to design the corresponding interface requires a clear sensor classification. Here we divide the sensor types into acoustic, electric, mechanical, optical, and electromagnetic and introduce its related physical sensing properties.

Typically, a sensor works in close collaboration with actuators and control unit to build the full cycle of an automated system, as illustrated in Figure 2. What a sensor measures will be interpreted by a logic unit, which is the decision making layer and leads to certain action triggered by it. An actuator acts the correct response according to the measured entity from the sensor.

In this survey, we only focus on the sensing part and portray all possible physical entities, which are commonly used to perform HAR. The miniaturization of sensing devices and the cheap production cost make smart sensing devices widespread in the smart home domain in an aim to simplify our everyday life. Voice assistants such as *Alexa, Siri, Cortana* and more [13] can listen to our voice command and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

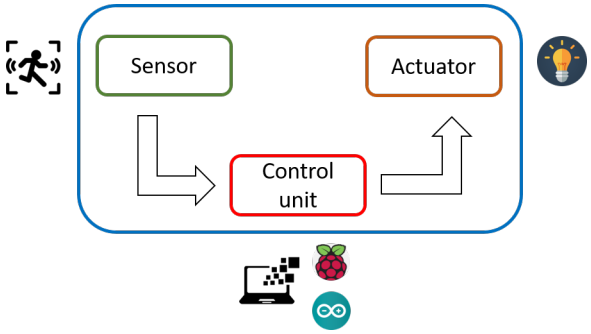

FIGURE 2. A sensor plays an essential part in an automated system. It senses certain properties of the environment and convert it to electric input feed to the central control unit. The control unit makes a decision based on the digital input data and makes the actuator act upon this decision.

control the lightening or other smart appliances. For human-centred designs, it requires to understand the human actions performed. Sensors can make the link between the human actions and the interpretation unit. The same human action can be measured by various sensor types, but the pool of actions is wide, which makes an action-based comparison more difficult. Therefore, in order to make a more easily comparison across sensors, we make the sensor classification based on the physical measures and provide related applications with this type of sensor used in the sub-domains of HAR, such as indoor localization [14]–[19], home behavior analysis [20]–[26], quantified-self [27]–[29], gestures, postures recognition [30]–[37] and sensing of physiological signals [32], [38]–[41].

Physical quantities, such as sound, light and pressure can be measured by acoustic sensors, optical sensors and pressure sensitive sensors. In the following sub-sections, we present some detailed works with regard to the sensor categorization given in Figure 1. The common structure for each sensor category is organized as follows:

1) introduce the physical sensing principle,
2) survey the most prominent research works that utilizes the questioned sensing category in activity recognition,
3) conclude and discuss the utilization of the sensing technology, including the advantages and disadvantages within the application domain,
4) a summery of the discussed works with a clear table-structured presentation of the main take-home-messages.

A. ACOUSTIC

Acoustic sensors can measure mechanical or acoustic waves traveled through certain materials. The transmission speed is affected by the different material properties over the propagation path in the transmission channel. Mechanical waves traveled through solid materials, can be detected by a surface acoustic sensor. Typical representatives of a surface acoustic sensor are built by piezo-electrical elements. These sensors are mostly operated in passive mode. Seismograph is a passive sensor, which could be used to measure the vibrations on the ground surface caused by a step signal. Passive sensors are compact, cost efficient, easy to fabricate, and have a high performance, among other advantages. However seismic sensors need a robust ground coupling to detect the vibrations traveled in the surface. The better the coupling, the better will be the signal to noise ratio of the received signals. Active acoustic sensor can measure sound waves transmitted through the air channel. These sensors can generate an electric signal, which will be converted to mechanical oscillation by using a membrane to set the air around the transducer into motion. This mechanical wave is modulated by the object or obstacles close to the sensor and the back reflection is sampled by an analogue digital converter (ADC) converting the echo modulation back to electric signal. In this subsection we will discuss three main categories of this sensing technology: active acoustic, surface acoustic, and ultrasonic sensors. This subsection will later include an overall discussion of the technology and a final conclusion.

1) Active acoustic sensors

Sound events such as clapping, coughing, laughing and yawning, besides natural speech languages carry additional information for perceptual aware systems. Schroeder [20] proposed using a microphone to detect four acoustic events (coughing, knocking, clapping and phone bell). Several signal processing steps and template matching from the frequency spectral domain are necessary to extract useful patterns to train the SVM classifier. Temko [21] focused on identifying 16 types of meeting room acoustic events, such as *chair moving, door slam, coughing, laughing, etc.*. Their sources of sound samples are acquired both from the public database, such as **RWCP** [42] , **ShATR** [43] database and the world wide web. However the class distributions are highly imbalanced, since the database with the targeted classes are mostly imbalanced.

One drawback of these acoustic sensor is, that these sound information collected by a microphone may also contain speech information and thus raise privacy issues. A viable solution is to use surface vibrations instead of sound signals.

2) Surface acoustic sensors

Pan [44] built a person identification system that utilized footstep induced structural vibration. The system sensed floor vibration caused by footstep without interrupting human activities. Gait analysis based on the characteristics of individual footstep was then exploited to achieve an identification accuracy of 83 %. By further incorporating a confidence level, the accuracy rate increased up to 96.5%. This was done by using only the most confident traces above certain threshold.

The signal to noise level of the received structural vibration signal is highly dependent on the sensor coupling to the ground and the surface materials. A sound coupling provides a higher signal to noise ratio. However, it is also possible to increase the detection accuracy by performing more signal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

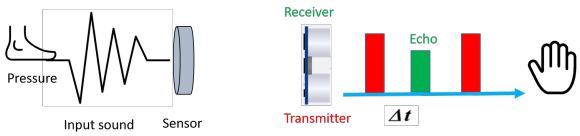

FIGURE 3. On the right side, the principle of a surface acoustic wave is depicted. Each footstep causes the surface to vibrate. This vibration can be measured by a microphone or seismograph. On the left side, a pulsed ultrasonic signal is depicted. Range information is unambiguous within two subsequent pulses.

processing on the input stage. Since these acoustic events contain high frequency component, neglecting the low frequency components of the vibration signal further concentrates the signal energy to a smaller frequency bands and thus further improves the signal to noise ratio. Mirshekari [45] managed to improve the localization accuracy of indoor footstep signals in this way. They were able to achieve an average localization error of less than 21 cm, resulting in an improvement of 13 times compared to the use of the raw input data.

Alwan [22] proposed a work to detect the fall event by leveraging a seismic sensor to catch the distinctive vibration characteristic of a fall event. Falls are most common among elders and are one of the leading cause of death for elders. The authors worked to distinguish patterns from dropping objects close to the sensor and simulated fall events from a Rescue Randy up to 20 feet away from the sensor. The detection of a fall event is based on the models according to the vibration patterns, such as frequency, amplitude, duration, and succession.

3) Ultrasonic Sensors

Ultrasonic sensors are active sensors, which actively transmit and receive signal to remotely perceive its environment. Ultrasonic spectrum starts from 20 kHz to 200 MHz, that is just above the human audible range. Ultrasonic sensing can be conducted in several classical forms. Acquiring distance information only, a pulsed sensor can be used to transmit high frequency pulsed signals and await for the reflected pulse bounced back off the measuring object. The operation frequency for most of the ultrasonic distance sensor are chosen to be 40 kHz. The time of flight, when the echo is registered by the ultrasonic receiver are correlated to the distance. The equation for calculating the object distance is thus $D = \frac{v_0 \cdot t}{2}$, where the speed of ultrasonic wave through the air is $v_0 = 340 \frac{m}{s}$ at a temperature of 20°C. Notice the 2 indicates the round-trip of the echo signal.

Acquiring motion information, such as the relative speed or moving direction, the Doppler measurement is required. To measure the quantity of Doppler broadening, a continuous signal of 40 kHz is emitted by the transmitter. The relative motion of a moving object is modulated above this carrier frequency. The amount of the Doppler in frequency shift can be calculated by using the Doppler equation, which then directly renders the information regarding speed and the sign is related to the direction of the relative movement.

TABLE 1. Table lists commonly used ultrasonic sensors used to build human activity recognition system in indoor spaces. Det. Range is the detection range, Field of view (FoV) and Op. Freq stands for operating frequency.

Sensor	Det. Range	FoV	Op. Freq
HC-SR04	4cm - 400cm	15° - 30°	40 kHz
LV-MaxSonar-EZ0	15.2 cm - 256 cm	33°	20 kHz

Indoor activities, especially activities of daily living, such as *standing, sitting and falling*, and *quantified-self* are the most popular use-cases for using ultrasonic sensors. Notably, for recognizing simple indoor activities, pulsed ultrasonic sensors are often used to measure distance towards the interacting object. Ghosh et al. [27], [46] mounted 4 HC-SR04 sensors to cover a square of 70 cm x 70 cm with a LV-MaxSonar-EZ0 in the middle to reduce the dead zone. Based on the distance profile, they used the support vector machine (SVM), k nearest neighbours (k-NN), and Decision Tree approaches to classify the targeted activities. The activities contain primitive activities such as *sitting, standing* and *fall*. Using Hidden Markov Model (HMM), they later extended their work to recognize these events for a group of multiple person [47] and the transitions of these primary states. Patel [48] targeted at a complete new set of activities of daily living including (*Nothing, Entered, Using Refrigerator, Used Refrigerator, Appeared near burner*, and *Using burner*) by applying Fusion of sensor networks consisting of Infrared Breakbeam Sensor, Ultrasonic sensor(HC-SR04) and Passive Infrared sensor(HC-SR501). The sensor specifications for the leveraged ultrasonic sensors are illustrated in Table 1. The operation frequency of the sensor, its field of view and the detection range are provided.

Physiological signals can likewise be detected by using a ultrasonic signal measuring the distance modulation of the chest movement during a respiration circle. Nandakumar [38] developed a contact-free sleep apnea detector with an off-the-shelf smartphone. They transformed the phone to an active sonar system by emitting linearly frequency modulated sound signals (from 18 kHz - 20 kHz) and extracted range information from the reflected echo signal caused by the chest movement. Hand gesture recognition task using a smartphone device is further targeted by the project *Dolphin* [30] and *FingerIO* [31]. Due to the limited detection range of a ultrasonic device, for close-range and fine-grained detection such as hand gesture and chest movement, a mobile application is more suitable than a fixed installation with a pulsed ultrasonic device.

4) Discussion

As stated in previously cited works in Subsection II-A, acoustic sensors, such as microphone, are mostly used to detect sound events, such as coughing, chair moving, door slam, transmitted through air. They are mostly used to infer sound-based events in private or public areas, such as a meeting room. Acoustic sound event is one of the most informative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

source besides natural speech to interpret a scene containing human beings and their interaction with the environment [20], [21]. These sensors don't require a solid coupling between the transmit medium and the sensor itself. However due to the nature of sound events, these sensors may raise privacy issues, since the general speech could be interpreted by the microphone.

Surface acoustic sensor measures the structural vibrations transmitted through solid materials. Since the production cost of these sensors are relatively low, they are often used to build distributed systems. It is power-efficient and its sparsity can further reduce the installation and computation costs. Applications built with this sensor type are mostly focusing on events causing vibrations on the ground surface, such as step signals [44], object dropping or fall events [22]. These events form a primitive set of activities of daily living in a household. However, sensors based on the structural vibration require a solid coupling between the sensor and the solid material. If the load on the ground surface is changed, the vibration intensity and the pattern previously extracted will also be deformed. These effects often lead to drops in the detection performance and require sensor calibration.

Ultrasonic sensors overcome both disadvantages, by transmitting and receiving high frequency signals to unobtrusively perceive its environment. The operation frequency is above the audible range of a human being and thus the audible spectrum can be excluded for processing. Opposed to surface vibration signals, no coupling to the ground is necessary. Integrated into the environment, it can sense object up to 2 m with a pulsed sensor operates at 40 kHz. Based on the distance profile, activities such as sitting, standing, and fall events can be recognized [47]. Operating in close range, it can detect fine-grained activities, such as gestures [30], [31] or even respiratory rate [38].

The usage of these sensor categories in the domain HAR are three-folds,

1) sound events detection related to natural sounds from activities of daily living with microphones,
2) surface vibration detection due to step signals with surface acoustic sensors,
3) dynamic activity recognition with ultrasonic sensors.

5) Take-Home Massage

One can notice that most works related to activities of daily living requires a network of this types of sensors. Due to the limited detection range of this sensor type, a full coverage of a room-scale requires multiple sensor fusion. Sound events, such as coughing, chair moving, or door slam can be detected by microphone arrays. Surface-bounded events, such as steps or falls are mostly measured by surface acoustic sensors. Fine-grained gestures or other delicate physiological signals require a close sensing range and high resolute senor system. For these applications, ultrasound sensors are preferred. An overview of the cited literature can be found in Table 2, where the previous works are introduced in terms of its application area, sensing device, processing algorithm, sensor behavioral, database and a concluding remark.

Work	Device	Area of Use	Algorithm	Active/Passive	Database	Remarks
[20]	Microphone	ADL	Self-organizing map (SOM)	Active	Private	Sound events in office
[21]	Microphone	ADL	SVM, GMM clustering	Active	RWCP, ShATR, Web	Sound events in public areas
[44]	Geophone	Identification, Gait Analysis	SVM, Multi-class C-SVC	Passive	Private	Hardware design required (Amplification circuit)
[45]	Geophone	Indoor Localization	TDOA	Passive	Private	Sensor networks required
[47]	Ultrasound Device	ADL	HMM	Active	Private	Echos, distance profiles, heterogeneous ultrasonic sensors
[38]	Ultrasound (Mobilephone)	Physiological signal	Peak detection in time series	Active	Private, clinically collected	Less power efficient, since continuous sensing
[30]	Ultrasound (Mobilephone)	In-air gesture recognition	Linear SVC	Active	Private	Multiple mobile devices
[31]	Ultrasound (Mobilephone)	Fine-grained finger tracking	Trajectory tracking	Active	Private	Enhanced interactive Display, OFDM signals, 2D finger position by leveraging two microphones

TABLE 2. Applications build on acoustic sensing.

B. ELECTRIC

The strength of an electric field is related to the amount of charge produced by an electrified object. When a detection

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

electrode is placed close to an electrified body, an electric charge proportional to the amplitude of the electric field is induced in the detection electrode. This physical effect is called electrostatic induction. The electric field can also be modified due to capacitive coupling with conductive materials or any other materials with a dielectric constant other than air. In the following, this subsection will introduce two main categories of this sensing technology: capacitive proximity sensing and electrostatic sensing with electric potential sensors. This subsection will later include an overall discussion of the technology and conclude with some final thoughts.

1) Capacitive Proximity Sensing

Capacitive measurement is based on electric field proximity sensing relying on the fact that an electric field is perturbed by the existence of a nearby conductive object, such as part of a human body. Therefore, this technology is often applied for remote sensing in the field of HAR. Capacitive sensing principle can be further divided into three operation modes, ranging from *loading modes, shunt mode*, and *transmit mode*, according to Smith [49]. In capacitive proximity sensing, the sensing category applies voltage to one side of the sensing electrode generating a constant electric field. The presence or motion of a conductive object close to the sensing electrode perturbs this electric field. The amount of the perturbation is directly correlated with the interactive item placed close by the sensing electrode. In Figure 4, the three operation modes of an active capacitive measurement are depicted. In transmit mode, the object acts as a transmitter and shortens the path of the electric field lines and amplifies the electric field. When the object is far from the receiver, the electric field weakens with $\frac{1}{r^2}$, since the object acts as a point source. Here r is the distance between the object and the receiver. While the distance decreases, the electric field weakens with $\frac{1}{r}$, as in this case the object acts as a parallel plane object to the receiver. In the shunt mode, electric field lines are partially occluded by the object and the electric field strength is weakened. In the loading mode, one can measure the displacement current from a transmitter electrode to a grounded body part. This mode is often used to get the relative distance from the sensing platform to the object.

Nowadays, capacitive technology can be found in almost every smartphone, tablet or touchscreen display. It is affordable and can detect the presence of fingers, hands or body movement with high accuracy. The project **Touché** by Sato [50] intended to enhance the touch interaction with capacitive sensing technique by leveraging the sweep frequency capacitive sensing technique. Conventional capacitive sensor operates at a certain frequency and can only detect touch interaction based on the amplitude modulation. By leveraging multiple frequencies, a more advanced profile can be built to include a variety of information, such as distinguishing between *not touching, touching, pinching*, and *grasping*.

Enhancing the touch modality, researchers design applications leveraging the proximity sensing ability of capacitive sensing. Proximity enables a more natural form of interaction compared to basic touch interactions. Braun [51] proposed a driver's seat enhanced with capacitive proximity sensing to detect a wide range of physiological parameters about the driver and his sitting postures for activity recognition in automobile applications. Identifying lying postures in bed, such as *supine, right lateral, prone, and left lateral* has been proposed by Lee [52] using the ECG signal of 12 capacitive coupled electrodes horizontally integrated into a bed cover. Rus [53] proposed similar lying posture recognition with mutual capacitance as sensor grid deployed under the mattress. These applications integrate the sensor electrodes into individual objects close to the sensing body.

Large-scale systems can also be built using capacitive sensing technique. Steinhage [14] proposed a smart floor using capacitive sensing that can be embedded under any non-conductive surfaces such as carpet or stone. Multiple features, such as person identification, persons path or trajectories tracking and fall detection are developed for this application. These features are especially useful to elderly care facilities. Similar work, **TileTrack** by Valtonen [15] based on transmit mode, measured the capacitance between multiple floor tiles and the receiver electrode to perform indoor 2D localization. The system with an operation frequency of 10 Hz can localize a standing human with an accuracy of 15 cm and a walking person within an error range of 41 cm.

Applications with capacitive sensing introduced so far are commonly focusing on static or stationary measurement such as sitting or lying postures and thus more stationary information are provided. Dynamic nature of the whole-body interaction and other remote activity recognition is sparsely exploited. This is partly due to the physical principle of static field measurement, but also a lack in this research direction.

2) Electrostatic Sensor

Electric potential sensor is an electrostatic sensor. Unlike capacitive sensing actively keeps a constant electric field to the sensing electrode, electrostatic sensor works with stationary electric charges. Electrostatic involves building up charge on the surface of objects due to contact with other surfaces. This charge induces an inverted charge on other opposite surface. Therefore electric potential sensors can be operated more power efficient due to the passive measurement of induced charges. However this induced charge is only noticeable, if the other surface has a high resistance to electrical flow and thus making the process of discharge remains long enough to be observed. Passive electric field measurement on the opposite is strongly dependent on the dynamic nature. The measurement solely based on body movement to generate body charges induced onto the sensing device. In case of electric field sensing, no constant electric field is applied on the sensing electrode. The sensing is merely based on the modulation of the existent ambient electric field caused by charge redistribution due to human motion. Thus, this sensing technology is strongly coupled to the ambient changes. The advantage of the electric potential sensors are light weight, large detection range, and high sensitivity. By using an ultra-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

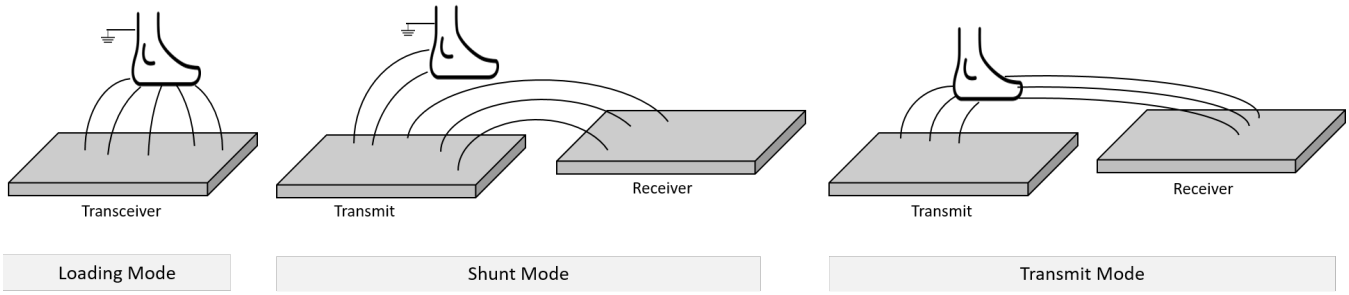

FIGURE 4. Figure depicts the three operation modes of an active capacitive measurement [49]. The electric field lines are depicted in black and are orthogonal to the surface.

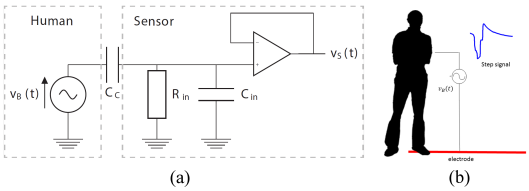

FIGURE 5. Figure shows the working principle of an electric potential sensor on the left and a typical step signal induced by the displacement current from the body motion [16].

high impedance sensor at the input stage, even the smallest displacement current caused by the body motion can be reliably measured. The working principle of such an electric potential sensor is viewed in Figure 5. The modulation of body induced current is illustrated by using an oscillating voltage source v_B and it is changing over time. The displacement current from the body motion is coupled between the body's surface and the sensor's metal surface with a capacitance C_c, which is typically in the order of $0.1 - 10\,\text{pF}$ [54]. This weak capacitive coupling requires a very high input impedance to reliably detect the minor displacement current generated by the body movement. Normally it is in the order of $10^{12} - 10^{15}\Omega$, to keep the output voltage v_s stable.

Prance [32] presented the ability of using an electric potential sensor to remotely detect physiological signals, such as the heart beat or respiration rate in a distance up to 40 cm from a seated subject. Rekimotor [55] built an enhanced game-pad using electrostatic potential sensing to allow whole-body input interactions such as (jumping, landing, foot lifting and foot touch) besides the general key press input modality. However, since the sensing principle is based on body charge modulation via body motions, most applications are focused on wearable designs, such as the work by [56]–[58].

Cohn [33] used a human body as an antenna for whole-body interaction in an indoor environment, by placing a miniature device on the body to collect the existent environmental "noises", such as AC power signal at 50 Hz or 60 Hz or other higher frequency signals from appliances and electronic devices. They leveraged the modulation of these electronic signals specific to differentiate activities caused by the body motion. They are able to sense 12 activities with an accuracy of 93 %.

Remote and embedded installation for this sensing technology have been developed mainly for indoor localization purposes, such as in the works [16], [17]. In the project **Platypus**, Grosse-Puppendahl [17] showed that by installing four ceiling mounted electric potential sensors covering an area of 2 m x 2.5 m, they were able to track people in a nearly empty office room around $16\,\text{m}^2$ with a mean localization error of below 16 cm. They found out that the electric pattern for each step for different person are distinctive within a short time window. Thus making use of the pattern recognition with handcrafted features by integrating priors from domain expert knowledge and based on some common features from literature regarding gait analysis, they are able to re-identify four users with an accuracy of 94 % and 30 users with an reduced accuracy of 75 %. Fu [16] deployed the measuring electrode in a grid-wise layout under a non-conductive floor covering to perform indoor localization. With a sensor electrode spacing of 20 cm and an system operation frequency of 10 Hz only, they achieved a mean localization error below 12.7 cm by leveraging a weighted mean position estimation method. The sensing area covers an area of 240 cm x 360 cm in a simulated living laboratory environment. The author stated, that this sensing technique is strongly dependent on the foot-wear of the users. The strength of the induced charge is strongly dependent on various aspects, such as the clothing, weather condition and foot wear, which makes the sensing system extremely susceptible to environment noise.

3) Discussion

According to the cited works in Subsection II-B, capacitive proximity sensing is commonly used to sense direct interaction modality such as touch interactions. It can also be applied to detect conductive objects up to 15-50 cm and thus enabling other applications expand the touch interaction. The sensing technique is well suited for measuring stationary objects, such as postures or other stationary information in close range to the sensing electrode. Thus for close range activities and stable detection, the active capacitive technique

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

is more preferable. Capacitive technology is widely used in touch screen technologies of the most current smart screen devices [59], such as smartphones, tablets or touch screens. Besides the basic touch interaction, the most common usage of capacitive proximity sensing is in static posture detection, such as sitting postures [51], lying postures [53] or falling events [14]. Large-scale installation is leveraged for indoor localization task [15] or reasoned to build system performing recognition of activities of daily living [14].

Technique of electrostatic sensing is used to better measuring the dynamic activities. In this case, the sensing is based on surface charge generation caused by movement. The produced surface charge induces an inverted charge on the opposite surface that is measured by a sensor with a relatively high input impedance. This type of sensor is light-weight, easy to deploy and power efficient, since no active electric field is generated and only the existent ambient electric field is exploited. This kind of sensor is applied in various use-cases ranging from sensing of physiological signals [32], to dynamic human activities [55], such as jumping, stepping or walking. Room-scale activity recognition [16], [17] with this kind of sensor is also possible. Even with a relatively low system operating frequency of only 10 Hz, an accurate indoor positioning system is achievable. Build upon this trajectories, researchers can easily conduct other extended researches such as gait analysis or behavioural analysis of the inhabitants. Combined with a reasoning system, Kirchbuchner [60] carried out predictions for early detection of dementia or other mental deceases based on these position contexts.

The usage of this type of sensors in the domain of HAR are two-folds:

1) close-range postures and stationary action detection with proximity capacitive sensors,
2) passive, far-range dynamic activity detection with electrostatic sensor.

4) Take-Home Message

Capacitive sensing technique is commonly used to detect stationary activities in close range, either direct touch or proximity up to 15 cm. Most common applications are finger touches, human postures or indoor localization. The resolution and detection range is directly related to the size, material and applied voltage on the sensing electrode. Capacitive sensor can produce ambiguous measurements. Placing a small object close-by results in the same measurement as a large object placed at a distant distance. This problem should be considered during the design phase. However the signals are consistent, such that it provides reproducible signals for same object under same measuring condition.

Electrostatic measurement of the electric potential sensor is commonly used to detect dynamic changes, such as body movements. The detection range of up to 2 m based on the hardware application is huge with respect to capacitive proximity measurement. However, the disadvantage of this sensing technique is that it is extremely susceptible to environment noise, which should be considered in the data processing stage. The signal patterns within a very short window is only reproducible, thus making it difficult to extract robust features directly from the signal pattern in time. The binary information of movement or non-movement can be leveraged to build precise indoor localization systems. Based upon the trajectories further applications can be researched. An overview of the cited literature can be found in Table 3, where the previous works are introduced in terms of its application area, sensing device, processing algorithm, sensor behavioral, database and a concluding remark.

C. MECHANICAL

Mechanical signal often indicates the force applied to a surface. The quantity of surface deformation is hence related to the impact of the interactive object. This can be expressed by the term $P = \frac{F}{A}$, where P is the pressure, F is the force applied in the normal direction to the surface and A represents the area of contact. The force induced deformation of the sensing surface, generates an electric signal, which is sampled by an analogue to digital converter to a quantitative measure. There have been many developments of pressure sensors in the past, which vary in terms of performance, technology, design and cost [61]. Its main application areas can be found in industrial monitoring, such as flow measurement or leakage detection [61]. In this subsection we will discuss two main categories of this sensing technology: resistive pressure sensing, and room-scaled pressure sensing with piezoelectric or fiber optical sensors. This subsection will discuss these two categories and later include an overall discussion of the technology and a final conclusion.

1) Resisitive Pressure Sensing

Applications for HAR with pressure-based sensing has been proposed in [23], [28], [62]. Xu et al. designed a *eCushion* to detect sitting postures. They used the resistive technology to measure the surface deformation by integrating fiber-based yarn which is coated with piezoelectric polymer [63]. The initial resistance of an unstressed surface is relatively high. With force applied to the textile, the intra-fiber distance is squeezed which makes the resistance to drop. By performing signal matching with dynamic time warping method, they achieved an overall recognition accuracy of 85.9 % for 7 sitting postures.

For quantified-self applications, Sundholm [28] developed a flexible textile equipped with a thin layer of conductive polymer fiber sheet consists of resistive pressure sensor matrix. The conductive sheet is positioned between 80 parallel stripes of conductive foil on each side (horizontal and vertical), resulting in a 80 cm $\times$ 80 cm sensor mat. The volume resistance of the fiber sheet changes locally, when the material is pressed. As output, a 80 $\times$ 80 pixel frame of the applied pressure can be sampled at 40 Hz. They recorded 10 exercises of 7 users, each exercise repeated 10 times over 2 different sessions per subject. These exercises included workouts such as *push-up, quadruped, abdominal crunch,*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

Work	Device	Area of Use	Algorithm	Active/Passive	Database	Remarks
[50]	Capacitive sensing	Smart Objects	Sweep Frequency Capacitive sensing	Active	Private	Exploratory Study
[51]	Capacitive loading mode	Sitting postures, physiological signal	Model-based approaches	Active	Private	Sparse sensor setup
[52]	Capacitive sensing	ADL	LDA, SVM, ANN	Active	Private, laboratory setup	12 capacitive coupled electrodes horizontally placed on the bed
[53]	Mutual capacitance	Lying postures	SVM	Active	Private, bedroom laboratory setup	Sensor grid integrated under the mattress
[14]	capacitive loading mode	Indoor Localization	Kalman Tracking	Active	Private	Sensors are modular
[15]	Capacitive transmit mode	Indoor Localization	Threshold in time series	Active	Private, constrained environment	Resolution is based on sensor size
[32]	Electrostatic Sensor	Physiological signal	Peak detection in time series	Passive	Private, medical center	Remote measure of heart beat
[55]	Electrostatic Sensor	Whole-body Gesture sensing	-	Passive	Private, constrained laboratory setup	Enhancing interaction for entertainment, Exploratory study
[33]	Electrostatic Sensor	Presence detection	Threshold	Passive	Private	Leverage the ambient electric power lines
[16]	Electrostatic Sensor	Indoor Positioning, Identification	SVM	Passive	Private, office room environment	Ceiling mounted
[17]	Electrostatic Sensor	Indoor Positioning	2D Positions	Passive	Private, living laboratory environment	Smart Floor installation

TABLE 3. Applications build on electric field sensing.

bridge, etc, and additional weight training such as *chest press with dumbbell* and *biceps curl with dumbbell*. An overall classification accuracy of 88.7 % for the person dependent and 82.5 % for the person independent case were achieved with a k-NN classifier. Template matching with dynamic time warping method was applied to count the repetitions. An average counting accuracy of 89.9 % across different subjects was achieved.

2) Room-scale Integrated Pressure Sensing

Other installed and embedded applications are focused on indoor positioning or detection of activities of daily living [64]. Integrating pressure sensors into furniture and floors in home environment, Lim [24] was able to recognize daily activities such as *meal, sleep, exertion, go-out*, and *rest* based on the object usage information. If anomalies in a healthy daily living style were detected, a warning sign was provided to care-givers or doctors without intrusion.

Similarly the **GravitySpace** [18] is an instrumented space used to track the user's location and their poses based on the physical imprints of the human force impact left on the sensing ground. Integrated with other modalities such as marker-based motion capture systems, audio-sensing equipment and video-sensing technology, Srinivasan [34] provided the pressure information as an additional input modality to enhance the application for interactive media usages.

Finally, the pressure can be measured not only with resistive technology, but also with fiber optics, as demonstrated in multiple works [65]–[67]. Feng [65] used floor pressure imaging for posture-based fall detection with fiber optic sensor grid-layout embedded under the floor space. People identification based on gait analysis problem has been targeted in the work by Qian [68]. Using a large area, high resolution, pressure sensing floor, they were able to provide 3D information of each footstep (containing the quantity of force and the 2D positional information). Applying the fisher linear discriminant classifier on the collected patterns from these 3D data points over time for each participant, they obtained an average recognition rate of 94 % and a false alarm rate of 3 % by using pair-wise footstep data from 10 subjects.

3) Discussion

Based on previously discussed works in Subsection II-C, we identified that pressure sensor arrays integrated into flexible textiles can be used in the applications for posture sensing or activity sensing. Build upon sitting posture recognition, researchers retrieve high-level contexts based on these primary information. Mota [69] tried to associate these naturally occurring postures and corresponding effective states relate to a child's interest level while performing a learning task on a computer. Features were extracted by leveraging a mixture of 4 Gaussian to express the force distribution on the back of a chair. A 3-layer feed forward network was used to train the classifier for nine postures and an overall accuracy of 87.6 % was achieved for testing on new subjects excluded from the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

training set. A set of independent Hidden Markov Models was used to link to three categories related to a child's level of interest. An overall performance of 82.3 % with posture sequences from known subjects and 76.5 % with unknown subjects were realized.

Textiles-based prototypes are flexible and easy to transport, however, they suffered from the problem of maintainability. Since the force is directly applied to the sensing surface, a flexible surface could be slightly deformed every time, it is used. Cheng [23] also noted that every time the Smart-Surface is installed and used, it was twisted slightly differently, which leads to a different default pressure distribution asserted by its own weight and folding. Further problems of textile sensors noted by Almassri [70] such as non-linearity, drifting and hysteresis could also influence the generality of the developed model for the target application.

Pressure sensors embedded under any floor covering or integrated into furniture as part of a distributed sensor networks can provide large-scale sensing in contrast to portable systems. They can be used to sense room-scaled indoor information such as location or other activities of daily living. Integrated into furniture or objects, theses objects can provide usage information to be accessed for smart home applications. Based on footstep force profiles, Orr [71] proposed a floor-based system to identify users in their everyday living and working environments. Creating user footstep models based on footstep profile features allowed them to achieve a recognition accuracy of 93 %. They've further shown, that the effect of footwear is negligible on recognition accuracy, in contrast to other sensor types, such as electrostatic sensing technique. Thus pressure sensors installed as a floor-based system enables a more robust and natural identification of users.

The usage of this sensor category in the domain of HAR are two-folds:

1) close-range posture, or action detection with flexible, resisitive textiles ,
2) room-scale sensing with either distributed pressure sensor networks or installed floor-based applications.

4) Take-Home Messages

Mechanical sensor works with pressure profiles caused by impact. Hence direct interaction is required. It is similar to active capacitive measurement by leveraging stationary force impact. Therefore, mechanical measurements are ideally used to measure postures or stationary activities. However, the proximity sensing would provide more information, including close range interaction as an additional input modality complementing the direct touch. Compared to passive electric field measurement, the foot-wear is negligible on the recognition accuracy for pressure sensing applications [71]. Thus this type of sensing technique is more error-resistant to the surrounding environmental noise, but bears the inherent problem of easier deformation. An overview of the cited literature can be found in Table 4, where the previous works are introduced in terms of its application area, sensing device, processing algorithm, sensor behavioral, database and a concluding remark.

D. OPTICAL

Optical sensors can quantify the intensity of light. Optical spectra cover a wide frequency range, from ultraviolet (280 nm - 360 nm) to visible (380 nm - 750 nm) to infrared (800 nm - 1000 nm). Invisible infrared light spectrum can be detected by infrared sensors, while visible light can be measured by the charge-coupled device (CCD) of a standard camera. In this subsection, we concentrate on the imaging ability of these optical sensing devices with the focus on HAR. We discuss three main categories of this sensing technology: visible imaging, depth imaging, and thermal infrared imaging. This subsection will later include an overall discussion of the technology and a final conclusion.

1) Visible Imaging

Vision-based HAR is probably one of the most well researched area in the field of computer vision, for enhancing the human machine interaction interface. Vision input compared to time series from sensor data provides more contextual information. From outdoor security applications [73], integrated with virtual reality techniques for entertainment purposes [74], monitoring and analysing sport activities [75]–[77], to medical applications [78], the demand on mature computer vision algorithms is growing.

Starting from segmentation [79] and recognition of human poses [80], towards continuous HAR [81], the full chain has been well studied. The most difficult part is to find feature representations in images to help developing robust human action modeling and thus improving the ability of algorithm to classify the correct activities. Unlike 2D image space, challenges in video sequence classification may include different appearances, shapes and poses in video frames over time and problems of occlusion from subsequent frames. From carefully handcrafted feature representations with expert prior knowledge [82]–[85], to the earlier stage of the deep learning era, a lot of efforts were made on developing robust models and generalized feature representations for accurate activity classification. Convolutional neural networks (CNN) like AlexNet [86], showed its superior ability to automatically extract useful feature representations from the underlying data structures. Other generative models, such as sparse autoencoders [87], and generative adversarial networks [88], are representatives of methods able to automatically learn the embedding representations of data.

Tran [89] studied a deep learning architecture for video action classification by extending a conventional 2D-CNN with a third convolution direction over time. The structure is called *C3D*. Their work demonstrated that this type of network is especially designed to extract features that model appearances and motion simultaneously. Input to the network is video clips of the dimension lxwxh, where l represents the number of frames per clip, wxh stands for the width and height of a frame and the output is the class probabilities of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

Work	Device	Area of Use	Algorithm	Active/Passive	Database	Remarks
[63]	Piezoelectric polymer	Sitting postures	DTW	Passive	Private	eCushion: challenge in form stability
[69]	Pressure Matrices	Sitting postures	MoG [72] + ANN	Passive	Private	relate naturally occurring postures to child's interest level
[28]	conductive polymer fiber	Fitness exercise recognition	k-NN	Passive	Private	Flexible sheet induces the problem of deformation
[64]	Weight sensitive floor	Indoor Positioning, Footstep signatures	HMM	Passive	Private	No EMC problems
[24]	Pressure sensors	ADL	Rule-based, attached to binary sensor states	Passive	Private	Integrated into furniture and floors in a home environment
[18]	Pressure sensing floor + touch sensitive furniture	Identify users, furniture and poses	Clustering for poses, Matching database for identification	Passive	Private	Updates in real-time and runs a physics engine to model the room above the surface
[34]	Pressure sensing floor	Dance movement	Pressure pattern from a footstep	Passive	Private	Multimodal sensing of a subject, including motion capture, audio and video sensing systems, high density of sensors required
[65]	Fiber Optics	Fall detection in smart home application	Pressure image and rule-based classification	Passive	Private	False positives due to missing dynamic links
[66]	Embedded hetero-core fiber optic nerve sensors	ADL in bathroom	Pressure mapping	Passive	Private	Activities in bathroom, prone to temperature fluctuations and electromagnetic interference
[68]	High resolution pressure sensing floor	Gait Analysis, People Identification	multi-classes FLD	Passive	Private	Strict assumption that features from different persons are linearly separable

TABLE 4. Applications build on mechanical sensing.

each activities. The network consists of several consecutive convolution and pooling layers to extract the high level appearance features and expand the field of view of the locally connected convolution features. However, it is to note that the first pooling layer only reduces the spatial dimension, but not the time dimension in order to preserve the temporal information further in the network. The performance was evaluated on three public available video databases: **Sports-1M** [90], **UCF101** [91] and **YUPENN** [92].

Another common design for video classification is the Two-Stream approach by Diba [93]. They showcased a similar model using two streams of 3D CNN. Such architectures are intended to solve the problem of insufficient training data as well as noise introduced by different view points, perspectives and variation in motions. The first branch, referred as the appearance stream, implemented the regular *C3D* network, while the second one, referred as the motion estimation stream, used optical flow as input. The features from the two streams were concatenated and feed to a softmax layer to infer the probability distribution of the classes. While testing on the UFC101 dataset [91], the two stream model outperformed the *C3D* network by 5 % with a 20 % decrease in processed frames per second. It confirmed the assumption that using optical flow helped the network recognize motion and complemented the appearance and spatio-temporal features learned by the standard *C3D*, however at the cost of decreased computational performance.

2) Depth imaging

The skeleton offers a more compact representation of the human body and enables simplified segmentation task and estimation of pose. Commercial products such as Microsoft Kinect makes visible images with depth information affordable. These devices can be used to capture human motions and provide the 3D coordinates (x, y, z) and the angle of the joints of the skeletons. The development of these skeletons over time in successive frames can be used to classify human activities of subjects within the measuring area. Compared to 2D images, the depth information facilitates the extraction of fore- and background.

Mostly, Microsoft Kinect is used to provide a depth channel in addition to visible channels. Official algorithm are provided to determine the skeletons and joint positions as features for various activity recognition tasks. Mettel [94] introduced a fall detection service using a single depth camera installed on the ceiling. Combining static and dynamic methods, a fall detection service was achieved by using a Microsoft Kinect. A random sample consensus (RANSAC) method was used to estimate the ground plane. Static detection verified whether the person was lying on the floor by tracking posture based on skeleton joint data. Dynamic detection examined whether a person is previously falling to the ground by thresholding the speed of the previous joint motion towards the ground plane. However, by placing only one single depth sensor in the room, the sensing area was restricted thus leading to performance degradation, when

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

the skeleton tracking was occluded by obstacles within the sensing area. Author proposed to use fusion of multiple installations to reduce false positives. Cippitelli [95] proposed an activity recognition framework to exploit skeleton data extracted by RGB-depth camera for recognizing activities relevant for assisted living. Their promoted use-case was to provide help to monitor aged people in home environments. Their main contribution was able to automatically extract key poses without a learning algorithm. The key poses were extracted using a clustering algorithm to assign each human posture to the most important posture for certain activity. The key poses were then concatenated to build a feature vector that was used for the multiclass SVM to perform activity classification. The proposed algorithm was evaluated on five public available databases (**KARD** [96], **CAD-60** [97], **UTKinect** [98], **Florence3D** [99], and **MSR Action3D** [100]) and showed promising results especially on a subset of basic activities designated from ambient assisted living scenarios.

GymCam [29] is a camera installation in a unconstrained environment, such as a university gym, which are then able to unobtrusively and simultaneously recognize, track and count fitness exercises performed by multiple persons. The promoted use-case is for quantified-self applications. It involves several computer vision tasks such as correctly segmenting exercises from other activities, recognizing and tracking users performing the exercise by following the trajectories of the interest points and counting the number of repetitions. Based on motion trajectories from key-points tracking using dense Optical Flow method, they were able to classify different activities from these features extracted by these motion trajectories. The repetition counting was based on template matching with an average trajectory of each exercise.

3) Thermal imaging

Images from visible light spectrum, such as visible images, may face a problem in object segmentation, if the appearances of the human subject, e.g. the color of the clothing is indistinguishable from the background. Thermal infrared imaging is resistant to this effect and can provide complementary advantages. Thermal cameras are passive sensors to measure infrared radiations emitted by any warm objects. Therefore, human motion can be easily detected from the background regardless of lighting conditions and appearance changes [101].

To use computer vision in pervasive health care is not new. Camera system installed in a living environment to detect activities of daily living is introduced in the work [25], [102], [103]. Person identification can be realized not only based on biometric trait such as face images, but may also based on soft biometric traits, such as gait pattern [104] or postures. To reduce the privacy concern regarding using cameras in domestic environments, low resolution thermal imaging method can be applied to achieve the detection of activities of daily living without revealing a wide range of private information. Shelke [26] used two low-resolution (4x16) and contact-free thermal imaging sensors (MLX90621) to classify four different activities such as *stand, sit on chair, sit on ground*, and *lay on ground*. For static activities, such as sitting on a chair or standing still, frame-wise classification were applied using conventional multiclass classifiers. Dynamic changes were observed via shape changing effect from consecutive frames due to motion relative to the sensor according to lens projection equation. The shape was detected by using connected component labeling approach [105] to group the corresponding pixels. The disadvantage of using the MLX90621 thermal sensor is its limited field of view (FOV). It has a 120° horizontal FOV, but only 25° vertical FOV. Therefore, a careful arrangement of sensor placement is required to achieve good performance.

Hevesi [106] leveraged a cheap (30USD), small, low power sensor array of 8x8 thermal sensors to unobtrusively and remotely detect a wide range of activities of daily living. The system can track people within the accuracy range below 1 m and detect the usage of electric appliances, such as toaster, water cooker or egg cooker. Basic activities, such as opening a refrigerator, the oven or taking a shower can also be detected. Due to the sparse sensor resolution by 8x8 pixels, the authors claimed that the system can be installed in the bathroom to recognize bathroom activities without invading privacy.

Kawashima [107] proposed a Deep Learning-based approach for action recognition method with an extremely low-resolution thermal image sequence. The hardware used is a grid of 16x16 far-infrared sensor array (Thermal sensor D6T-1616L by OMRON Corp.) mounted on the ceiling (around 220 cm above the floor) of a room. They focused on recognizing daily activities, such as walking, sitting down, standing up etc. and abnormal activities (e.g. falling down). The authors combined feature extraction method based on shallow CNN structure (consisting of only 3layers), with a sequence layer based on long short term memory (LSTM) for extracting spatio-temporal representation. With a frame rate of 10 fps, the overall accuracy for the targeted activity classes were 85.75 %. Data collection consisted of sequences from day and night times. The superiority against visible light is that the night vision for thermal imaging can make a "falling down" action in the dark visible in contrast to a total black visual input in visible light spectrum.

4) Discussion

In accordance with the cited works in Subsection II-D, camera systems provide richer information compared to other non optical sensors accompanied with the cost of more computation efforts. Recent advances made in computer vision domains ignite more interests in this field. Especially, faster progress was made in object detection and localization with algorithms such as YOLO [108] to faster YOLO [109], and Fast R-CNN [110] to Faster R-CNN [111]. The tendency is to work on faster algorithms, which can be embedded on hardware with limited resources. The development from

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

semantic segmentation with Mask R-CNN [112] and Eye-MMS [113] to instance segmentation with DeepMask [114] also allows for more precise information retrieval for separating instances from the same class. Video sequence processing with C3D network or attention network for sequence input [115] make activity recognition in complex scene possible. Despite the advanced algorithms, challenges such as occlusion, change of appearance and prone to illumination changes, are only partly resolved for camera systems.

To reduce the negative effect of illumination changes, additional channel of depth can be integrated. The information of depth is used to resolve the ambiguity in two dimensional image space. Commercial products from Microsoft and Intel make depth camera accessible for researchers to conduct experiments in the field of computer vision with depth channel. Microsoft Kinect automatically comes with the joint positions of the skeleton model. The skeleton representation is more sparse and compact, thus enabling more efficient processing on embedded hardware entities. Skeleton-based processing is commonly applied for human action recognition. Based on handcrafted features extracted and well-designed classifiers, human skeleton can be used to extract spatial structure and temporal dynamics specific from human actions. Lately, research interests shift to consider end-to-end learning to avoid handcrafted features and model construction with prior knowledge. Du [116] proposed an approach based on hierarchical recurrent neural network to learn representations of skeleton poses hierarchically fused from sub-nets to automatically form action models fitted for the separate action classes. Skeleton-based approaches for HAR to build assisting system for elderly monitoring was introduced in [94], [95]. Activities of daily living, such as sitting, standing, walking, and falling are the most often targeted classes.

Thermal infrared imaging is another sensing form operating with near to far infrared light spectra. The operating wavelengths enable the system to observe radiations emitted by objects with a temperature above zero. Therefore it facilitates the segmentation process from human object to background. Night vision capability of infrared sensors even enable action recognition in the dark opposed to image data from visible light spectrum. It also enables the reconstruction of visible-like images from thermal captures [117]. Infrared sensor arrays used in the cited works are mostly sparse and can be applied to reduce the resolution to protect users privacy. Sensor array of 4x16, 8x8 or 16x16 pixels are used. These installations are often applied in home environments to build systems for tracking and evaluating activities of daily living.

The usage of these sensor categories in the domain HAR are three-folds,

1) camera-based action recognition in public areas,
2) depth-based action recognition and tracking on embedded hardware platforms,
3) low resolution thermal infrared imaging in home environments to build ambient assisted living systems.

5) Take-Home Message

Action recognition in computer vision can be performed on images, videos, or life streams. Each of the target domain bears its own challenges. Image covers only one instance in time and thus context can be missing if the decision is based on only one single image. Action recognition in video requires more complex network architecture to integrate the time component. Real-time assessment of human activities can enable robots to operate intelligently in interaction with humans. Part of these challenges have been already solved by the modern deep learning methods. By using 3D network structures or sequence modelling methods, the aspect of time is considered. Knowledge distillation [118] or network pruning [119] can decrease the model capacity and make real-time assessment possible.

Despite the rapid development in computer vision, one of the biggest drawbacks of camera based solutions is the low user acceptance in private sectors, as cameras typically raise concerns about privacy [120]. Therefore, either using depth channel or using thermal imaging can help resolve some of the mentioned challenges for visible spectral input. An overview of the cited literature can be found in Table 5, where the previous works are introduced in terms of its application area, sensing device, processing algorithm, sensor behavioral, database and a concluding remark.

E. RADIATION

Radiation, in the form of electromagnetic waves, works with high frequency electric field modulations. Common custom radar in the automotive domain operates at a typical frequency of 24 GHz [121] and 76 GHz [122]. On the other hand, according to WiFi standard 802.11n [123], domestic WIFI frequency bands operate at 5 GHz for close range and 2.4 GHz for far range. The operating frequency of 2.4 GHz grants for better penetration through solid objects and thus provides a wide coverage of WIFI signals. In the following, this subsection will introduce two main categories of electromagnetic sensors: radar sensors and WiFi sensors. This subsection will later include an overall discussion of the technology and conclude with some final thoughts.

Sensor devices generating a high frequent electromagnetic field, such as a radar, can operate in two different modes, in continuous wave (CW) mode and frequency modulated continuous wave (FMCW) mode. In the CW mode, only relative speed toward the receiver can be measured, while the FMCW can also provide distance information with the time beacon information encoded in the start frequency. In Figure 6, the two operation modes of radar are visualized. For continuous wave radar depicted on the left, if the transceiver and the distant object are both stationary, the received signal is not modulated. If the distant object is moving with a speed of v relative to the receiver, then a positive or negative Doppler shift can be measured for an approaching or departing object. Since there is no timing information available, only the relative speed represented by a Doppler profile can be extracted from the continuous signal. For frequency

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

Work	Device	Area of Use	Algorithm	Active/Passive	Database	Remarks
[73]	Video surveillance	Crowd behavior recognition	KLT tracking, crowd motion vector and clustering	Active	Private	same type of event may be modelled slightly different from one scenario to another and many models are to be defined.
[104]	RGB image	ADL, Gait Analysis	Model-based approach	Active	KTH, Weizmann	Model-based approach is more constrained and less variable
[25]	RGB video sequence	ADL, multi targets detection	Blob detection based on Optical flow, Color-based tracking	Active	Private, Smart Laboratory setup	Vision as an additional input modality to enhance the ability of existing no-vision sensors
[102]	Distributed smart camera network	Fall detection, localization	Threshold on aspect ratio, mapping image coordinates to world coordinates	Active	Private, staged living room condition	not prone to false positives due to the missing dynamic links
[103]	Multiple cameras	Tracking of interacting and occluded human motion	Bayesian belief network for multi cameras fusion, Kalman filter	Active	Private, simulated home environment	can not handle self-occlusion
[26]	Thermal imaging	ADL	LR, NB, SVM, DT, RF, ANN	Passive	Private	Restricted field of view for thermal imaging sensors
[107]	Thermal imaging	ADL	CNN+LSTM	Passive	private, day and night vision of thermal inputs	16x16 grid resolution to preserve privacy
[29]	Logitech C922 camera	Gym exercises recognition and tracking	Optical flow, feature extraction of motion trajectories, MLP neural networks for classification	Active	Private, Carnegie Mellon University's varsity gym	Limitations to handle intra-class variability
[83]	Skeleton image sequences	Activity recognition	CNN + attention	Active	NTU RGB+D, SBU Kinect Interaction dataset UFC 101, Kinects dataset	can handle noisy poses
[84]	RGB+Optical Flow	Activity recognition	Two-stream network	Active	HMDB51, UFC101	Combine handcrafted features with deep learnt features
[85]	Video streams	Event detection in soccer broadcasts	Tracking with particle filters, SVM	-	FIFA World Cup 2006	Highlights and tactic detection, game analysis on broadcast videos

TABLE 5. Applications build on optical sensing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

modulated continuous wave case depicted on the right, based on the time shift of the received signal with respect to the transmitted signal, a distance profile can be generated in addition to the speed information.

WiFi sensing also depends on similar sensing protocols. However, it can further access the channel state information to infer HAR. Channel state information (CSI) describes the channel property between the transmitter and the receiver. Radio signal from the transmitter can travel directly to the receiver (LOS), but may also be scattered by objects or reflected by walls and ceiling before reaching the receiver. CSI can be represented by the channel transmission matrix, describing these different effects, such as fading, scattering, and multi-path fading, by the physical environment between transmitter and receiver. Common WiFi systems use Orthogonal Frequency-Division Multiplexing (OFDM) [124] to divide the wide spectrum band into around 30 non-overlapping subcarriers. In this case, CSI contains complex values, which represents the channel properties of each subcarrier. Take a WiFi channel in the 2.4 GHz band with multiple inputs and multiple outputs (MIMO) mode, containing 3 Transmitter and 3 Receiver antennas, the CSI Tool can capture 30 OFDM subcarriers, resulting in 3x3x30 CSI data points in each received packet for processing [125] at each time instance. By collecting these CSI data points over time, we can build a CSI profile used to capture the changes in the physical environment. A moving object such as a human being in the receiving path can affect the channel response and be measured on the receiver side.

1) Radar sensors

We start introducing the radar sensing, which has the advantages of insensitivity to environment conditions and robustness in different weather conditions. This makes radar applications in HAR a suitable candidate. It can transmit signal through walls, thus no direct line of sight is needed compared to vision-based systems. Using millimeter waves, the resolution is so fine that it can even detect the smallest finger movement in the order of sub-millimeter. Motion sensing with **Soli** [35], a tiny radar chip to detect and recognize hand gestures developed by Google has now been commercialized and integrated into Google's new smartphone *Pixel 4* [126]. Rahman [39] proposed yet another contact-free measurement of respiration rate by leveraging the phase shift in Doppler radar signal caused by the chest movement and allow person identification based on the subtle body kinematics of six individuals. A 2.4 GHz quadrature system is used to reduce the DC offset to allow more amplification and thus increasing the dynamic range of detection.

Seifert [127] used radar-based applications to perform unobtrusive person identification based on In-home gait analysis. A K-band radar was used to collect data from four test subjects. K-band operates in the frequency range of 18–26,5 GHz, the radar used here is at 24 GHz. In their proposed work, different walking styles were further clustered into five different gait classes including *normal, pathological* and *assisted walks*. By leveraging the radar micro-Doppler signatures, an average identification accuracy of 93.8 % was achieved across the classes and a classification rate of 98.5 % was achieved for a single gait class. A performance drop to 80 % accuracy was expected for unknown individuals. Features from both the spectrogram and cadence velocity diagram were extracted based on prior expert knowledge. A simple classifier using nearest neighbour (NN) approach was applied to the handcrafted features condensed by the principle component analysis technique (PCA).

Liu [128] leveraged a dual Doppler radar system for fall detection operating at 5.8 GHz covering a detection range of 6 m. They used the Mel-frequency cepstral coefficients (MFCC) [129] to extract features from the Doppler signatures caused by different activities. The decision of fall/non-fall detection was then based on fusion of multiple trained classifiers output.

Deep learning technique has also found its way to radar signal processing as in computer vision applications. Most of these methods are directly applied on time-frequency spectrum (spectrogram). Similar to computer vision tasks, where CNN is applied on images to extract features for object recognition, CNN can analogously be used on spectrum images to extract spectral patterns resulting from specific activities. Kim [130] proposed a deep convolutional neural network architecture for human detection and activity classification based on Doppler radar operating at 7.25 GHz for outdoor and 2.4 GHz for indoor activity recognition with direct line of sight. This network jointly learned the feature representations and classification in one single network based on the raw Doppler spectrum. Activity classes included *running, walking, walking while holding a stick, crawling, boxing while moving forward, boxing while stand in place*, and *sitting still*.

Similar to time series for natural language processing, recurrent neural networks (RNN) can support the decision making stage of activity classification by considering the time aspect of the signal progress. However for radar images, a 2D-CNN layer is often applied prior to the RNN layer in order to extract robust features from the time-frequency spectrogram. The follow up work of using **Soli**, a customized, miniaturized radar chip to resolve sub-millimeter gesture motions, showed such a network structure in [36]. Their network consisted of two stages including the representation learning stage by using a CNN network, followed by the dynamic sequence modelling stage of using a long short term memory (LSTM) network prior to the classification stage with a Softmax layer. They achieved a per frame accuracy of 79 % and a per sequence accuracy of 88 % on a set of 11 hand gestures across 10 different users.

Ultra-wide band (UWB) is a radio technology that is used at short-range, high-bandwidth communications. It has been widely used in radar imaging domain. Compared to CW radars, it exceeds in terms of range resolution. Compared to FMCW radars, the UWB transmission is able to send very short pulses mitigating the multi-path inference prob-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

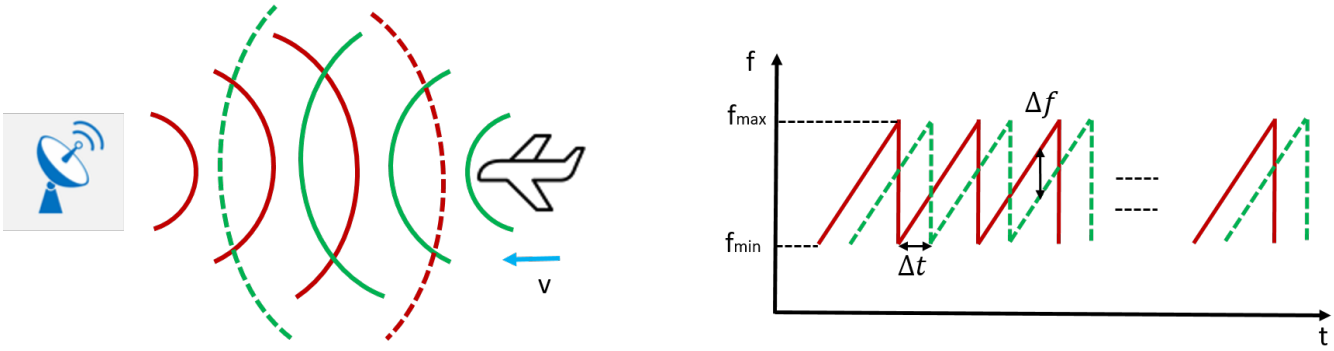

FIGURE 6. Figure shows the two operation modes of a radar. On the left side, if both transmitter and object are stationary, then no Doppler shift can be measured. In case of a moving object with a speed v, a positive or a negative Doppler broadening will be measured relative to the motion direction to the transmitter. On the right side, the **FMCW** mode is depicted. Based on the time shift, the distance of the object towards the transmitter can be calculated using the time of flight. The transmitter signal is shown in red, while the receiver signal is depicted in green.

lem. UWB operates commonly in the frequency spectrum of 3.1 GHz to 10.6 GHz, a broad frequency bandwidth of more than 500 MHz and a very short pulse duration of (<1 ns) [131]. This property makes the signal hard to detect and thus it is immune from detection, jamming, and interference. Lai [132] leveraged a UWB random noise radar to characterize human activities and through-wall imaging. So-far, the use-cases for radar imaging with UWB radars are mostly concentrated for military purposes or served for law-enforcement. They can be used in the search and rescue operations. Ding [133] conducted a thorough investigation on a large number of motion types based on an UBW radar system. They clustered different motions into two main categories of motion, including in situ motions and non-in situ motions. They leveraged physical empirical features for classifying in situ motions, such as standing, bowing, squatting etc., and the PCA-based feature extractions for inferring non-in situ motions, such as walking, jogging, jumping forward and falling forward. They reported a final classification accuracy of up to 94.4 % and 95.3 % for in situ motions and non-in situ motions, respectively. They claimed that their proposed method could be used in smart homes and senior care domains.

Radar is good for dynamic activity recognition, because of its robustness and its high resolution, as they operate at several gigahertz range, but it comes with the price of high power consumption and complex hardware design. WiFi devices are much more power efficient compared to radar sensors, if one can accept the comparable lower resolution.

2) WIFI sensors

Most radar comes with a high specialization and integration between hardware and software packages. In order to fulfill certain task specification, a separation between the software layer and hardware layer are often needed. This makes embedded radar packages difficult to be specialized for a broad range of applications in the HAR domain. Therefore, researcher tried to find a replacement which has similar physical behaviours, but are easier to modify and access. Researchers state that the channel state information (CSI) from a WiFi signal can be leveraged to passive and unobtrusively monitor the presence or motion of a human being. Popular application of using wireless devices for indoor localization based on WiFi fingerprint is quite common, such as introduced in [19]. When a person comes in the way between a WiFi transmitter and receiver, it changes the received signal strength (RSS) transmitted to the receiver. This modulated RSS profile can be used to extract useful information with respect to activity classification.

WiGest [37] is a ubiquitous wifi-based gesture recognition system to sense in-air hand gestures by leveraging the modulation in WiFi signal strength around a mobile device, such as a consumer smartphone. Based on three basic primitives, such as *approaching, removing*, and *holding above the device*, they were able to composite high level gestures without training for gesture recognition. With only one Access Point, they were able to detect the basic gestures with an accuracy of 87.5 %. To further include three Access Points, they were able to increase the accuracy to 96 %. Adding preambles as the start of a intended gesture, additionally improved the recognition accuracy and reduced the interference from multi-user scenario.

Accessing only the CSI of a WiFi signal, Zeng [40] built an application to monitor human respiration even when the target is far away from the WiFi transceiver pair. Common WiFi based application needs the object to be close to the transceiver, because the attenuation of radio frequency (RF) signal operating at 2.4 GHz is around 6 dB for a solid wood door with 1.75 inches and almost 9 dB for an interior hollow wall with a depth of 6 inches [41]. Instead of working directly with the raw CSI signal, they leveraged the CSI signals from two transmitters to cancel out the environmental noises and benefited from the phase information of the cleaned signal.

WifiU [134] is a gait recognition system that uses an commercial off-the-shelf (COTS) WiFi devices to leverage the channel state information to capture fine-grained gait

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

patterns for person identification. In contrast to expensive Doppler radars, the channel state information can provide similar information such as motion from echos caused by back-scattering from different body parts. *WifiU* consisted of a router and a receiver to collect the modulated CSI due to human motions. A WiFi device sends continuously signals to its environment which are scattered by moving objects, such as a human within the transmission path. The scattered signals are then received by a laptop. A PCA-based technique is used to reduce the environmental noise signals by extracting the principle components from the correlated CSI signals. The true movement resulted in dominant components within each sub-carriers and uncorrelated noise components were suppressed by using the PCA method. After applying PCA, the time echo was still composed of reflections from various body parts. The decomposition of such a time signal was performed by using a frequency-time spectrum (STFT) method. The reason of using Fourier transformation on the time signal was that different body part moves at different speed resulting in different Doppler shifts. The main goal was thus to transform the received CSI signals to the Doppler spectrum similar to other radar-based applications with high fidelity to extract Doppler motion information. Higher speed corresponded to higher Doppler shift and vice versa. Feature extraction was then performed on the cleaned Doppler shift profiles.

WiSee [135] is another application for sensing whole-body gesture recognition by leveraging wireless signals in an office environment or a two-bedroom apartment. Pu leveraged the frequency-time Doppler shift profile from various body parts while performing specific tasks, to achieve an recognition accuracy of 94 % on a set of nine gestures such as *push, pull, circle, dodge, drag, punch, strike, kick*, and *bowling*. Adib [41], developed by MIT researchers, showed various interesting use-cases by leveraging COTS WiFi devices. They were able to count persons, locate their relative positions, measure vital signs such as respiration rate and heart beat rate even from an adjacent room or behind closed doors. By treating a moving human as a moving antenna array, they were able to build an inverse synthetic aperture radar (ISAR) technique to enable radar-like vision. They can scan the movement of the human in time by only using one single antenna.

WiFi-based activity recognition utilizes existing wireless transceiver infrastructure in the environment to measure activity induced WiFi signal variations. Compared to radar-based applications, WiFi application is more power efficient and preserves user's privacy, since no physical sensing module is required except the already existing WiFi communication route.

3) Discussion

As reported by the cited works in Subsection II-E, electromagnetic sensors are resistant to different weather conditions or environmental noise at certain operating frequencies. In contrast to optical vision-based system, high frequency electromagnetic waves do not require a direct line of sight and can even penetrate through walls. In addition their robustness, safety, and reliability make them perfect to serve as an effective device for contact-free and ubiquitous motion monitoring of objects in the surrounding.

Due to its robustness against extreme weather conditions and large detection range, radar-based applications are already widespread in automotive sector for environment sensing and perception. Operating in the sector of HAR, the operating frequency and the transmit power should be reduced to adapt to indoor applications. Most use-cases work with radar sensors operated around 5.8 GHz, 7.25 GHz or 24 GHz. Human motions such as gait [127] or other whole-body interactions [130] can be leveraged to developed human-centered smart home appliances. Even sub-centimeter resolution of finger gestures can be observed with the specialized and miniaturized radar device *Soli* [35] integrated into a smartwatch or smartphone device.

For close range radar applications, UWB radar are often applied. Its advantages include low power consumption and is more secure due to extreme short pulses, high transmission rate , noise resistant due to ultra wide-band. Related to its superior physical properties, UWB can be used to perform exact indoor localization. The short duration of UWB pulses make them robust to multipath effects, since the identification of the main path from other multipath signals is more evident and thus allowing a more precise detection of the time of flight [136]. Through the wall object imaging [132] is another useful task for UWB imaging radar, especially in situations where a direct line-of-sight is not possible. For example, it can be used in rescue operations or finding tracked person in a collapsed buildings.

WiFi application is more power efficient compared to general radar applications or UWB radars. Most WiFi-based applications work with modified WiFi access points. Compared to integrated hardware and software solutions of most radar applications, it is easier to modify the WiFi access points to adopt to specific tasks designed for HAR. Common applications build with modified WiFi devices are targeted at tracking and recognition of indoor activities. Close range applications include near device in-air hand gesture recognition [37]. Room-scaled applications are commonly focusing on indoor localization [19] and tracking of human [41]. Based on Doppler profiles, whole-body gestures [135] can be targeted even when the sensor is placed behind the walls. Applied for localization tasks, the maximum detection range is up to 250 m outdoor and 35 m indoor [136].

The usage of these sensor categories in the domain HAR is three-folds:

1) dynamic fine-grained whole-body activity recognition with radar-based sensors,
2) close-range fine grained activity recognition and imaging with UWB radar,
3) more power efficient whole-body activity recognition with WiFi signals.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

4) Take-Home Message

Radar applications are mostly used in outdoor environments with large operation frequencies, large detection range and high operating power such as environment sensing and perception of a vehicle on a motorway. Applications in indoor environments in case of human activity classification need to operate with lower frequencies and lower operating power. Most use-cases work with radar sensors operated around 5.8 GHz, 7.25 GHz, or 24 GHz. In case of CW radar or FMCW radar, a continuous signal is transmitted all the time, making these applications less power efficient. For close range detection and sensing UWB radars are often applied due to its preferable physical properties.

However, most radar hardware are difficult to build. Commercial radar solutions have hardware and software packages strictly coupled such that an easy modification of radar software adapting to specific use-case is not accessible. One alternative is to use the channel state information of a commercial WiFi system. WiFi signals are easy to access and more power efficient compared to radar based applications, but operate at much narrower operation frequency bandwidth of only 20 MHz compared to 1.79 GHz for a FMCW radar, resulting in lower time resolution than radar applications. An overview of the cited literature can be found in Table 6, where the previous works are introduced in terms of their application area, sensing device, processing algorithm, sensor behavioral, database and a concluding remark.

F. OTHERS SENSOR TYPES

Other physical quantities, such as temperature, chemical composition, and magnetic field modulation can be measured by dedicated sensors. However these sensors are not often used as a single sensing entity in the field of HAR [137]. Human activity is complex and it requires to capture information from multi-sensor networks to infer the correct actions [138]. Variables such as temperature may add low level information to the process of activity reasoning, however, information fusion is needed to integrate the data in the high level decision making process. Magnet sensors can be placed on furniture, drawers, or doors to provide binary information when users directly interact with these objects [139]. Temperature, light, pressure, humidity, or CO2 sensors are all components that can be used to build a wireless sensor network for smart home systems [140]. ZigBee [141], for example, is used as a low cost, low power, less complex wireless communication standard to connect such sensor nodes with the main processing unit in a smart home system.

Applications integrating magnetic sensors into MEMS placed in initial measurement units (IMUs) are used for pose and acceleration measurement, mostly in wearable devices, such as smartphones, smartwatches, or other miniaturized on-body devices. Altun [142] used five body worn sensors placed on the chest, the arms, and the legs to classify daily and sports activities of eight subjects. Each sensor integrates a triaxial gyroscope, a triaxial accelerometer and a triaxial magnetometer. Combining feature dimension reduction techniques, such as PCA and sequential forward feature selection (SFFS) methods with Bayesian decision making classifier, they were able to balance between a high correct classification rate with relatively low computational cost with regard to real-time application.

Fusion multiple sensor categories to infer human actions is advantageous, because different sensor type provides different context (place, time, situation, etc). To ease the decision making process, a richer context is beneficial. Even combining multiple sensors from the same category, such as combining multiple acceleration-based sensors on the human body can increase the recognition accuracy of complex human activities. Maurer [143] investigated the classification accuracy of wearable sensor on different body position. Results demonstrated that the sensor placement strongly affected the recognition performance and could lead to misclassification if not properly placed.

Bao [144] revealed that two out of five bi-axial accelerometers were enough to recognize a set of 20 activities including ambulation and daily activities such as *scrubbing, vacuuming, watching TV,* and *working at the PC*. By only using the sensors on the hip and wrist as a sub set of all locations, the accuracy only decreased around 5 %. An increased accuracy of 25 % was achieved over the best performing single acceleration sensor. The fusion was performed on the feature-level by concatenating extracted raw features from the acceleration data time windows. However activities such as *stretching, scrubbing, riding escalator* and *riding elevator* were often confused. To overcome this issue, they required additional sensor modalities. Heart rate data can for example reveal the intensity of physical activities and GPS location data can provide the information whether the individual is at home or at work, and thus add a probability measure to certain set of activities.

Chernbumroong [145] proposed a multisensor framework for activity recognition with genetic algorithm (GA) [146] to determine the fusion weights of the multisensor platform. The multisensor platform consists of accelerometer, temperature sensor, and an altimeter on a CC430F6137 Microcontroller with MSP430 CPU from Texas Instruments. Pressure sensor, gyroscope, barometer, and light sensor are integrated on Gadgeteer FEZ Cerberus board. In addition, a heart rate monitor is fixed on the chest with a chest strap. The sensor fusion was performed both on the feature and decision-level (classification-level). To compensate for sensors that are less dependant in making decisions by themselves, such as altimeter and temperature due to their low-level context, these outputs were fused at feature-level to provide a richer context. The used feature selection was based on the feature importance. On the decision-level fusion, the outputs of multiple classifiers were fused using GA method to fine-tune the fusion weight parameters. The sum fusion on the decision-level improved the classification accuracy from 96.9662 % of the best single classifier to 97.3096 %. In 98 % of the experiment trials, the GA fusion method outperforms the one best single classifier.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

Work	Device	Area of Use	Algorithm	Active/Passive	Database	Remarks
[35]	Millimeter wave radar	Fine-grained gestures	Bayesian inference	Active	Private	Motion resolution within sub-millimeter accuracy
[39]	CW radar	Physiological signal	Majority voting, k-NN	Active	Private, laboratory setup	Restricted study with 6 participants only
[128]	Dual Doppler radar	Fall detection	k-NN, SVM, Bayes	Active	Private	Aggregation by fuzzy integral, data acquisition from a trained actor only
[130]	Doppler radar with IQ-Demodulator	Military activities	DCNN	Active	Private, constrained laboratory setup	End-to-End training directly on the raw micro-Doppler spectrum
[127]	K-band radar	Identification based on Gait	Feature extraction with Prior PCA + Nearest Neighbour	Active	Private Office environment	Fuse features from both spectrogram and cadence velocity diagram
[36]	Millimeter wave radar	Fine-grained finger gestures	CNN + LSTM	Active	Private	End-to-end training
[19]	Wireless	Indoor Positioning	RSS, TOA, TDOA, FDOA	Active	-	Survey
[37]	WiFi RSSI	Hand gestures	Template matching, correlation	Active	Private, living laboratory condition	No additional modification to available wireless equipment or any extra sensors
[40]	WiFi-based	Physiological signal	Maximize breathing-to-noise ratio, template matching	Active	Private, typical home environment	Signal processing makes it possbile to detect respiractory signal even through walls
[41]	WiFi MIMO, USRP software radios	Single, multiple person tracking through walls	MUSIC, to disentangle correlated, super-imposed signals	Active	Private, office building	Moving target detection in the CSI channels
[134]	Commercial WiFi device	Fine-grained gait patterns	Peak detection on the upper contour of torso movement from the spectrogram	Active	Private	Limitations on path and direction of the experimental setup, single person use-case
[135]	USRP-N210s	Whole-body gestures	Pattern matching with predefined templates	Active	Private, office and two bedroom apartment	Preamble required, pre-defined template requires prior knowledge

TABLE 6. Applications build on electromagnetic sensing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

Similar fusion methods are reported in the field of multi-biometric fusion [147], where other methods that take advantage of multi-decision coherence [148], variations in information source trust [149], or relative relation between confidence levels in multiple sources [150], can be mapped into the multi-sensor fusion in HAR applications.

Therefore, the context provided by one sensor category is limited. To infer complex human actions, richer context is required which can only be done by fusion of different sensor modalities. Integrating additional sensor or sensor categories can boost classification accuracy by achieving the following gains as reported in [151] and initially defined by Bellot et al. [152]:

1) Accuracy gain: accuracy of decisions and representations after the fusion process is improved. Noise and errors are reduced in comparison to single source information.
2) Completeness gain: the information after the fusion process is less redundant and more complete.
3) Representation gain: the information after fusion is more granular compared to each of the single fused sources.
4) Certainty gain: the belief in the fused information is increased.

III. POPULAR DATABASES

In this section, we introduce several publicly available databases for the task of HAR, which are commonly used as baseline for researchers. They can be divided – based on our discussed sensor categories – into three groups: the single non-vision sensor category, the multiple sensor category, and the vision-based datasets. An overview of these databases can be found in Table 7.

1) Datasets using only one single sensor category

In the **Intel Research Lab** dataset [158], the authors used the RFID technology to recognize routine morning activities. They installed 60 RFID tags in the kitchen on objects touched by the user during a practice trial. The user wore two gloves built by Intel Research Seattle to detect that an object has been touched. However, unlike bar-codes, RFID tags can not specify uniquely which instances of objects have been touched, rather that some objects have been touched.

The UCI daily and sport dataset (**DSADS**) [154] consists of 8 subjects performing 19 different activities by wearing acceleration sensors on 5 body parts. Besides the more stationary classes such as *sitting, standing*, or *lying*, they also include dynamic exercises such as *ascending and descending stairs*, and *exercising on a stepper or a cross trainer*. However this dataset is only restricted to on-body wearable devices, where each wearable has a gyroscope, an accelerometer and a magnetometer.

The **PAMAP2** dataset [155] aims at physical activities such as *walking, cycling, playing soccer*, etc. It composes of 9 subjects performing 18 activities with 3 inertial measurement units and a heart rate monitor. Compared to **DSDAS**, this dataset fused another sensor category by integrating the heart rate monitor to provide additional information. As stated in [145], fusion of several sensor modalities can provide richer context to improve the performance of recognition on more complex human actions.

2) Datasets using multiple sensor categories

Previous cited databases are either ubiquitous or wearable. However they only used one single sensing category and thus the provided context was limited. Thus, other databases also use a composite of object sensors and ambient sensors to further incorporate more sensing modalities. The MIT **PLIA** dataset [157] are collected in a real experimental environment of 1000 sq.ft. apartment. *PlaceLab* is a new live-in laboratory for studying ubiquitous technologies in home settings. Approximately 214 sensors such as state sensors, accelerometer, camera, ambient sensors and object sensors were installed in the laboratory environment. During a 4-hour period, 89 activities are manually labeled from the collected sensor data.

The **CMU-MMAC** dataset [163] is another database leveraging multi-modal sensor data input for detecting tasks involving cooking and food preparing. Modalities collected are video, audio, motion capture, IMUs and two wearable devices. The dataset consists of five subjects cooking five recipes, in average 15 minutes/recipe. In this database, people and objects are visually instrumented and thus making the videos less realistic. The limited number of only 5 dishes with very similar ingredients and tools lead to restricted data variances.

The **MPII Cooking Activities Dataset** [164] tried to close this gap of limited and constrained variations by providing a large database with more realistic, fine-grained activities. The database contains 65 different cooking activities performed by 12 participants. Instead of recording individual activity, the participants were asked to perform actions in sequence and recorded by video to reflect a more realistic behavior.

The **TUM Kitchen** dataset [139] aims to provide a comprehensive collection of sensory input data, to serve researchers in the field of marker-less human motion capture, segmentation and activity recognition. It collects of video data with four fixed overhead cameras, RFID tag readings and magnetic sensors detecting when a door or drawer is opened. All four subjects perform the same high level activity of setting a table. The dataset was constructed such, that it tackled challenges which is not covered in other available datasets. Those challenges are such as inter-class variability, change of human silhouette while interacting with objects, human performing several actions in parallel, occlusion by furniture, and subtle actions.

The **Amsterdam** dataset [156] records the in house activity data of a 26 year old man, living alone in a three-room apartment monitored by 14 state change sensors placed in different locations, such as on doors, cupboards, refrigerators, and a toilet flush sensor. Authors stated that the upgrade ability of their system is advantageous compared to other

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

No.	Database	Sensor	Task	#Instances	#Subject	#Activity	Remarks
1	OPPORTUNITY [153]	Hybrid sensor networks	ADL	2551	4	4	large scale deployments of heterogeneous networked sensor systems
2	UCI DSADS [154]	multiple on-body accelerometers	ADL and Sport	9120	8	19	dataset is restricted to wearable
3	PAMAP2 [155]	Accelerometer plus HR-Monitor	Physical Activity	3850505	9	18	enhanced basic context with additional sensor category
4	Amsterdam [156]	Hybrid sensor network	ADL	245	1	7	upgrade-ability of sensor nodes, missing generalizability with dataset from one person
5	MIT PLIA [157]	Hybrid sensor networks	ADL	-	-	89	live-in laboratory
6	Intel Research Lab [158]	RFID tags	ADL	-	1	11	Single sensing category with binary information
7	Mocap [159]	optical marker-based	Motion clips	2605	144	6	very clean and detailed motion capture data
8	Sports-1M [90]	vision-based	Sport clips	1,133,158	-	487	under unconstrained environment and contain complex activities
9	UCF101 [91]	vision-basec	Sport clips	13000	-	101	under unconstrained environment and contain complex activities
10	KTH [160]	vision-based	sport clips	2391	25	6	simple, isolated actions
11	MSRDailyActivity3D [161]	vision-based	ADL	960	10	16	action dataset of depth sequences noisy 3D joint positions if person stands close to a background object (sofa)
12	TUM Kitchen [139]	Hybrid sensor networks	setting a table	36666	4	10	tackles large pool of activites
13	URADL [162]	vision-based	ADL	50000	5	17	contains high resolution video sequences of complex actions, each containing only one specific task
14	CMU-MMAC [163]	Hybrid sensor networks	cooking, food preparing	-	25	5	limited variations due to small number of dishes
15	MPII Cooking Activities Dataset [164]	vision-based	ADL	881755	12	65	fine-grained actions

TABLE 7. Popular Time Series Databases for Activity Recognition Tasks.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

datasets [157] where sensors should be installed during the contraction time for intended locations especially build for research purposes. They claimed that if people are living in an unfamiliar environment, the action collected are not representative. Their solution is to leverage sensor network consists of wireless network nodes to which simple off-the-shelf sensors can be integrated. In such a way, they can easily upgrade the user's living environment with wireless sensor networks. However, the dataset of only one person is limiting the results of its general validity.

The **Opportunity** database [153] is often used as a baseline dataset for HAR collected from wearable, object, and ambient sensors. It consists of 4 users performing activities of daily living in an indoor environment. They deployed a wide range of 72 sensors of 10 different modalities in 15 wireless and wired networked sensor systems. The authors claimed that most existing datasets [156], [157] are not sufficient enough to investigate opportunistic activity recognition, where a large amount of sensors is required not only in the environment, but also on the body and in objects.

3) Vision-based dataset

Image based databases for HAR tasks are not rare. Datasets with constrained whole-body interactions, or the target on outdoor sport activities are provided in [160], [161], [165]. The **KTH** database [160] currently contains 2391 sequences and is collected under four different scenarios with 25 person performing six different activities, including *walking, jogging, running, boxing, hand waving*, and *hand clapping*. This dataset only includes simple, isolated actions in staged data. No complex actions or multiple person case are targeted in this dataset. The data acquisition process is performed under constrained scenarios. The task of simple action recognition can be considered as "solved", since most techniques already report nearly perfect results [166], [167].

Compared to the **KTH** database, the **URADL** dataset [162] contains high resolution video sequences of complex actions. It includes 10 different activities such as *answer phone, chop banana, drink water, eat snack, look up in phone book, etc.*, and are collected with high-resolution videos installed overhead. Even some classes are very similar, thus introducing more inter-class similarity, the scenes per video are constrained and each containing only one specific task.

Fully unconstrained datasets in the wild are collected in [90], [91]. The **Sports-1M** is a database [90] collected from the web, containing 1,133,158 video URLs, which has been automatically annotated with 487 labels. Also, the **UCF101** dataset [91] consists of 101 action classes, over 13k clips and 27 hours of video data. This dataset contains user uploaded activities with unconstrained data collection process, containing camera motion and cluttered background. The unconstrained setting poses a challenging task for precise action recognition with computer vision methods.

Research in vision-based action recognition has made a lot of progress with the advances in deep learning and computer vision methods. Researchers moved on from recognizing simple, constrained actions to more complex actions or interactions with multiple person under unconstrained environments. Therefore, such databases containing unconstrained conditions and multiple complex scenarios, are considered to be more useful in this regard.

4) Discussion

Datasets with only single sensor category provide limited context and thus making it difficult to tackle more complex human actions. Therefore, databases composited of multiple sensor modalities or even the same sensing modality on multiple locations help to solve more naturalistic and complex human actions. Common hybrid databases use composition of sensor modalities with low level information, such as state sensors, acceleration sensors, temperature sensors, and RFID tags. Image-based or video-based databases can provide rich context, however, often suffer from the problem of occlusion and privacy issues. If taken in private sectors, users may feel observed and thus do not act naturally or not representative of their usual behaviours.

Capacitive sensors or radar sensors can provide complex high-level information without violate the privacy. However, most of radar application did not make their databases public. Ideally, a composition of these high-level information reasoned from capacitive, radar or WiFi sensors can be fused with low-level binary sensors instead of using vision-based systems, especially given the privacy concerns connected to vision-based sensors. The ability of these sensor to observe activities even through walls, makes them strong against occlusion and the line-of-sight problem. High frequency radar devices could resolve fine-grained action within sub-centimeter range and thus making the recognition of fine-grained and more complex actions possible.

IV. EVALUATION METRICS

HAR can be treated as a pattern recognition problem, with the patterns related to specific actions. A list of the commonly used classifiers in the literature separated according to its categories can be found in Table 8. The most used classifiers and action detection methods in HAR can be divided in three large categories,

- Generative models: A generative model is a probability based method to learn the statistical distribution of the underlying data distribution. Generative model is able to create new samples based on the learnt statistics of the data distribution.
- Deterministic models: Deterministic models are static classifiers trying to learn the hidden feature representations from the labeled training data. Discriminative model is intended to determine the membership of each sample to a certain class.
- Others: Other methods include non-parametric methods. Non parametric methods make no assumption of statistic distribution from the given data. They try to draw conclusions about the data from data with similar patterns.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

Novel methods like the compressed sensing based HAR classification methods are currently drawing more and more attentions. These methods work with sparse representation and benefit from correlations in data to increase the processing speed and enable designers to place applications on devices with limited computing power. Examples of that are the works [168] and [169] where the authors explored compressed sensing based HAR classification methods and achieved satisfactory results.

Evaluation metrics are needed to compare different approaches and performances of action recognition systems. Though, the most metrics are defined for binary classification problem, they can be easily extended to fit multiclass classification problem. In this case, the multiclass problem can be divided into several binary classification problems. In Table 9, the most used evaluation metrics are given. As reported by Ward et al. [180], a valid methodology for performance evaluation should fulfil two main criteria:

1) The metric should be objective and unambiguous. The outcome should not dependent on random assumption or parameters.
2) It should provide a quantitative measure to give a hint to the strengths and weakness of the system or method.

V. DISCUSSION

Physical sensors are limited by its hardware and software characteristics. In the following, we discuss the hardware features related to the introduced sensor categories. We then identify some general challenges while performing software processing for these sensor categories.

A. SENSOR HARDWARE CHARACTERISTICS

Each sensor technology has its own advantages and disadvantages, limiting its use in various specific target applications. To select the appropriate sensor category or a combination of sensor categories for specific task is a design choice based on various aspects. To better compare sensor categories to each other, standardized sensor specifications can be taken into considerations. In Table 10, we introduce some feature matrix denoting capabilities required for a certain rating. We grade the features into five categories, ranging from ($--$, $-$, o, $+$, to $++$). The scoring is based on the research papers collected in this manuscript and sensor specifications found from sensor data sheets. Some features depend on the use-cases and the form factor of sensor categories. Power efficiency for instance, is thus strongly dependent on the underlying system setup and not solely on the sensor technology. Similarly, the sensitivity is also a feature strongly related to how the sensor is applied in the specific system setup. Some of the discussed features are not quantitatively evaluated in previous works or are not measurable as a scalar. Therefore, we introduce our ranking for these features as a relative measure based on the description of the user experience. These features are, such as calibration complexity, weather dependency, form stability, electric noise coupling and occlusion. According to the assessment criteria presented in Table 10, the different sensor categories is graded in Table 11.

Acoustic sensors can work both contact-based or contact free according to the specific task requirements. Contact-free sensors, such as microphones can classify human activities by leveraging acoustic events, but may raise privacy issues similar to a vision-based imaging system. Ultrasonic sensors on the other hand work in close range up to 5 m even in darkness. Thus, it is invariant to illumination changes and weather resistant. However, since these systems are active, the power efficiency is worse than other electric field measurement sensors, such as capacitance sensor or electric potential sensors.

Active capacitive sensing can work up to 15 cm in close range, but it is more noise prone, as noisy detection in far range can not be resolved by the sensing system. Passive electric field measurement is purely passive and is sensitive up to 2 m in range. Electrostatic sensors work purely passively and are thus more power efficient. As the sensor is extremely sensitive to the ambient electric field, the system is prone to electric appliances or ambient power lines. This requires hardware filters in the electronics design phase to reduce the power-lines coupling around 50 Hz.

Mechanical sensors respond to direct touch and are thus less susceptible towards power-lines or other ambient noise. Pressure signals are reproducible when the same force is applied, unlike electrostatic sensor depending strongly on the varying ambient electric field. On the other hand, mechanical sensors are more susceptible towards form stability. Especially, pressure sensors integrated into flexible textiles are prone to deformation. Deformation may easily break the pressure sensor or lead to performance degradation.

Vision-based systems are one of the hottest research areas for HAR. With techniques based on deep learning and large amount of online image resources, researchers are able to build robust segmentation and action detection algorithms. But the hardware limitation of the imaging system in visible spectrum, such as incapability of illumination resistance, occlusion, and change in object appearances over time, makes vision-based system still a challenging topic.

Electromagnetic sensors are more resilient to environment coupling than any other treated sensor categories. They are robust against weather or climate changes operating at certain frequencies. They can cope with changing illumination or even occlusion cases, because signals can even penetrate through walls at certain operating frequencies. The hardware is designed such that the life span is long and the form stability is high. To reduce the power consumption of radar-based devices, a modified WiFi access point can be leveraged to perform similar dynamic activity recognition tasks. Common commercial radar sensors closely connect hardware and software solutions together, such that an easy modification of the software with respect to a custom specification is not possible. WiFi devices, on the contrary, can be easily modified to gain access to the channel state information. The resolution accuracy of WiFi devices is lower in comparison

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

TABLE 8. Some of the most popular algorithms used for action recognition and classifiers in HAR along with examples of the works that utilized them.

Category	Abbreviation	Algorithm name	Source
Generative	HMM	Hidden Markov model	[64], [170]
Generative	DBN	Deep Belief Network	[171], [172]
Generative	NB	Naive Bayesian	[170], [173]
Generative	GMM	Gaussian Mixture Model	[21], [174]
Generative	denoise AE	denoise Autoencoder	[170]
Discriminative	SVM	Support Vector Machine	[27], [46]
Discriminative	CRF	Conditional Random Field	[175]
Discriminative	LR	Logistic Regression	[176]
Discriminative	CNN	Convolutional Neural Network	[36], [89], [130]
Discriminative	LSTM	Long Short Term Memory	[36]
Discriminative	RNN	Recurrent Neural Network	[36]
Discriminative	ANN	Artifical Neural Network	[26], [177]
Others	k-NN	k-Nearest Neighbours	[27], [46], [71], [170]
Others	DT	Decision Tree	[178]
Others	LDA	Linear Discriminant Analysis	[179]
Others	DTW	Dynamic Time Warping	[28], [62]
Others	FLD	Fisher Linear Discriminant	[68]
Others	compressed sensing	Compressed Sensing	[168], [169]

TABLE 9. Some evaluation metrics commonly used in HAR along with examples of the works that utilized them.

Application	Metric	Definition	Source
Identification	top-1 accuracy	number of positives for all genuine match within the top 1 returned candidate list	[134]
Identification	top-k accuracy	number of positives for all genuine match within the top k returned candidate list	[134]
Classification	AUC	area under the receiver operating characteristic (ROC) curve indicating the performance of a binary classifier system	[178]
Classification	Accuracy	ratio between number of correct predictions to total number of predictions	[36], [62]
Classification	Recall	proportion of actual positives to correctly identified	[181]
Classification	Precision	proportion of positive identifications to actually correct	[181]
Classification	F1-score	a weighted measure between recall and precision	[182]
Classification	Confusion Matrix	indication of how well each class preforms and gets confused with	[37], [39]

TABLE 10. Feature matrix denoting capabilities required for a certain rating. List of Features are Resolution (res), Update Rate (upd), Detection Range(det), Unobtrusiveness (unob), Processing Complexity(proc), Calibration Complexity (calco), Sensitivity (sens), Life span(ls), Weather Dependency (wi), Form stability (fs), Electric noise coupling (enc), Occlusion (occ), Power Efficiency (pe).

feature	$--$	$-$	o	$+$	$++$
res	<8 m	<100 cm	<30 cm	<20 cm	<10 cm
upd	<1 Hz	<10 Hz	25 Hz	>50 Hz	>100 Hz
det	touch	<1 m	<5 m	<20 m	>20 m
unob	open large system	open small system	hidden, large exposure	hidden, noticeable exposure	invisible
proc	single sensor CPU	10+ sensors CPU	single sensor, embedded chip	10+ sensors by single chip	no further processing
calco	very hard	hard	normal	easy	very easy
sens	insensitive	less sensitive	normal	sensitive	highly sensitive
ls	<3 years	<5 years	5 years	>10 years	>15 years
wi	dependent	less robust	neutral	robust	invariant
fs	deformable	less stable	stable	robust	rigid
enc	highly sensitive	sensitive	normal	less sensitive	insensitive
occ	fatal	prone	neutral	stable	invariant
pe	>1000 mW	>750 mW	300 mW	<220 mW	<25 mW

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

TABLE 11. Benchmark sensor system with respect to feature matrix given in Table 10.

Sensor	res	upd	det	unob	proc	calco	sens	ls	wi	fs	enc	occ	pe
Microphone	+	++	++	-	-	−−	++	-	−−	++	++	++	+
Vibration Sensor	+	+	o	+	−−	-	+	+	++	o	−−	-	++
Ultrasonic Sensor	+	++	o	o	-	o	+	++	-	++	−−	-	o
Piezo-electric	++	-	−−	++	-	++	−−	−−	++	-	++	o	++
Fiber-optical	++	-	−−	++	-	++	−−	+	++	-	++	o	-
Capacitive	-	+	-	++	o	+	+	++	-	+	−−	+	+
Electrostatic	+	+	o	++	+	+	++	++	−−	++	−−	++	++
RGB cameras	++	o	+	-	o	-	o	−−	-	++	++	−−	−
Infrared imaging	-	+	o	-	+	o	-	+	+	++	++	−−	+
Radar	-	+	++	+	o	o	++	++	+	++	++	++	−−
WiFi	-	+	++	+	o	o	+	++	o	++	++	++	−−

to high frequent radar applications, but with much reduced power consumption.

Therefore, how to choose the appropriate sensor category is strongly dependent on the design choice. According to range, obtrusiveness, robustness, and resolution, multiple sensor categories can be leveraged. Complementary sensor categories can be fused to provide richer context information to adapt to more complex human actions.

B. SENSOR SOFTWARE CHARACTERISTICS

Regarding the software processing step, data-driven models extremely rely on the underlying data distribution. The performance is thus directly related to the data availability and data acquisition process. We identified some data-related challenges and software design issues encountered in the domain of HAR with sensor data. The following challenges are mainly divided into

- computation time,
- data acquisition process,
- database availability,
- data distribution,
- data augmentation ability,
- the intra-class and inter-class variability.

These aspects are considered to be important while designing a robust model to perform HAR with sensor data. In general, the process of data acquisition and the labeling task for HAR system are tedious and expensive. Extensive manual labelling and expert knowledge are required. While image-based data are easy to acquire from the web or public databases, other non-visual data is less frequently available. There are several officially available databases with the focus on activity recognition for image or video data as introduced in section III. Images can be easily augmented using simple computer vision techniques, such as rotation, zooming, random cropping or applying noise filters to increase the amount of the training data. But it is not the case for time series. Time series are special, because the sequential information encoded in the time series can not be easily ignored. During the research phase, we identified that most of the applications with non-visual sensors collected their own database within a moderate test study and have not made it publicly available. Therefore, either unsupervised machine learning techniques should be applied to cope with the problem of missing labels, or shared database as benchmarks especially for time series data is desirable.

TABLE 12. Research fields in human activity recognition with some common applications.

Research field	Applications
Quantified-self	[27]–[29]
Home behavior analysis (ADL)	[20]–[26]
Gesture recognition	[30]–[37]
Video Surveillance/ Analysis	[73], [75]–[77]
Gait analysis	[44], [45], [68], [104], [128], [134]
Posture estimation	[51]–[53], [62], [63], [65], [183]
Physiological signal sensing	[32], [38]–[41]
Indoor Positioning	[14]–[19]

VI. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

HAR is the key to enable human-centered application and natural interaction in a smart environment. To solve this challenge, the ability to learn the knowledge about human activity from raw sensor inputs is of vital importance. Therefore, we revised various research activities in this area and defined a number of sensor categories to perform this task. In Table 12, sensor-driven applications with respect to the target domain in the area of HAR are depicted.

According to the surveyed most prominent research works in this manuscript, we summarize in Table 13 the different sensor category used for certain applications in the domain of HAR. Given an illustration like this, it is easy to identify missing application domains and provide some ideas for future research directions.

We further identify some challenges to be faced in this research field of action recognition with the previously introduced sensor categories. The main challenges can be categorized as follows:

1) Real-time detection, instead of offline processing: This requires smaller models, which can be applied on embedded devices with less computation powers. The capacity of the models should still be big enough to catch the underlying data representation.
2) Online-learning: Most of the machine learning models trained today are based on a fixed amount of training

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

TABLE 13. It illustrates the sensor categories used for each application in the domain of human activity recognition. We can easily identify missing application domains with certain types of sensor categories and future research directions.

Application Domains	Acoustic	Ultrasonic	Capacitive	Electrostatic	Pressure	Camera	Radar	WiFi
Activity of daily living	x	x	x		x	x	x	x
Physiological signals		x	x	x			x	
Quantified-self					x	x	x	x
Postures Detection			x		x	x		
Gestures Detection		x	x			x	x	
Gait Analysis	x		x	x				x
Indoor Localization	x	x	x	x	x	x		x

data and thus do not generalize well on new data. The ability to cope with new, unseen data, without the need to train the model again is thus a new requirement on the current model. The model should possess the ability of progressive learning.

3) Transfer learning and cross domain adaptation: The process of labeling HAR tasks is tedious and expensive. Therefore, if we can transfer knowledge from existing domain into a new domain with only less or mostly unlabeled data, it will save a lot of time and human resource of labeling.
4) Target the problem of inter-class and intra-class variability: Human motion is highly complex and possess a high degree of freedom. This can be expressed with the term user-diversity. Therefore, to design a robust model to cope with every possible situations, researchers should first target the problem of reducing the intra-class variability and increase the inter-class variability.

With the recent advances in computer vision and deep learning, we are convinced that the above mentioned challenges can be efficiently targeted and solved. Different sensor categories provide its own advances and disadvantages. During the design phase, researchers should weigh their choices according to the design goals required. Fusion of complementary sensor categories can sometimes also increase the performance and provide additional information to overcome their individual limitations.

REFERENCES

[1] Gerald A Dienel. In memoriam louis sokoloff, m.d. 1921–2015. Journal of Cerebral Blood Flow & Metabolism, 36(2):278–280, 2016. PMID: 26661214.
[2] Karl-Franz Reinhart and Matthias Illing. Automotive Sensor Market, chapter 2, pages 5–19. John Wiley & Sons, Ltd, 2005.
[3] Richard M White. A sensor classification scheme. IEEE Transactions on ultrasonics, ferroelectrics, and frequency control, 34(2):124–126, 1987.
[4] Oscar D. Lara and Miguel A. Labrador. A Survey on Human Activity Recognition using Wearable Sensors. IEEE Communications Surveys & Tutorials, 15(3):1192–1209, 2013.
[5] Salwa Slim, Ayman Atia, Marwa M.A., and Mostafa-Sami Mostafa. Survey on human activity recognition based on acceleration data. International Journal of Advanced Computer Science and Applications, 10, 01 2019.
[6] Xinyu Li, Yuan He, and Xiaojun Jing. A survey of deep learning-based human activity recognition in radar. Remote Sensing, 11(9):1068, 2019.
[7] Shuangquan Wang and Gang Zhou. A review on radio based activity recognition. Digital Communications and Networks, 1(1):20–29, 2015.
[8] Shian-Ru Ke, Hoang Le Uyen Thuc, Yong-Jin Lee, Jenq-Neng Hwang, Jang-Hee Yoo, and Kyoung-Ho Choi. A review on video-based human activity recognition. Computers, 2(2):88–131, 2013.
[9] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, and Lisha Hu. Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119:3–11, 2019.
[10] Sreenivasan Ramasamy Ramamurthy and Nirmalya Roy. Recent trends in machine learning for human activity recognition—a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4):e1254, 2018.
[11] Diane Cook, Kyle D Feuz, and Narayanan C Krishnan. Transfer learning for activity recognition: A survey. Knowledge and information systems, 36(3):537–556, 2013.
[12] Zawar Hussain, Michael Sheng, and Wei Emma Zhang. Different approaches for human activity recognition: A survey. arXiv preprint arXiv:1906.05074, 2019.
[13] Matthew B Hoy. Alexa, siri, cortana, and more: an introduction to voice assistants. Medical reference services quarterly, 37(1):81–88, 2018.
[14] M. Sousa, A. Techmer, A. Steinhage, C. Lauterbach, and P. Lukowicz. Human tracking and identification using a sensitive floor and wearable accelerometers. In 2013 IEEE International Conference on Pervasive Computing and Communications (PerCom), pages 166–171, March 2013.
[15] M. Valtonen, J. Maentausta, and J. Vanhala. Tiletrack: Capacitive human tracking using floor tiles. In 2009 IEEE International Conference on Pervasive Computing and Communications, pages 1–10, March 2009.
[16] Biying Fu, Florian Kirchbuchner, Julian von Wilmsdorff, Tobias Grosse-Puppendahl, Andreas Braun, and Arjan Kuijper. Performing indoor localization with electric potential sensing. Journal of Ambient Intelligence and Humanized Computing, 06 2018.
[17] Tobias Grosse-Puppendahl, Xavier Dellangnol, Christian Hatzfeld, Biying Fu, Mario Kupnik, Arjan Kuijper, Matthias Hastall, James Scott, and Marco Gruteser. Platypus - indoor localization and identification through sensing electric potential changes in human bodies. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '16, New York, NY, USA, 2016. ACM.
[18] Alan Bränzel, Christian Holz, Daniel Hoffmann, Dominik Schmidt, Marius Knaust, Patrick Lühne, René Meusel, Stephan Richter, and Patrick Baudisch. Gravityspace: Tracking users and their poses in a smart room using a pressure-sensing floor. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, page 725–734, New York, NY, USA, 2013. Association for Computing Machinery.
[19] Ashraf Tahat, Georges Kaddoum, Siamak Yousefi, Shahrokh Valaee, and Francois Gagnon. A look at the recent wireless positioning techniques with a focus on algorithms for moving receivers. IEEE Access, 4:6652–6680, 2016.
[20] Jens Schroeder, Stefan Wabnik, Peter W. J. van Hengel, and Stefan Goetze. Detection and Classification of Acoustic Events for In-Home Care, pages 181–195. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
[21] Andrey Temko and Climent Nadeu. Classification of acoustic events using svm-based clustering schemes. Pattern Recogn., 39(4):682–694, April 2006.
[22] Majd Alwan, P.J. Rajendran, S. Kell, David Mack, S. Dalal, M. Wolfe, and Robin Felder. A smart and passive floor-vibration based fall detector for elderly. volume 1, pages 1003 – 1007, 01 2006.
[23] Jingyuan Cheng, Mathias Sundholm, Bo Zhou, Marco Hirsch, and Paul Lukowicz. Smart-surface: Large scale textile pressure sensors arrays for activity recognition. Pervasive and Mobile Computing, 30:97–112, 2016.
[24] Joon-Ho Lim, Hyunchul Jang, Jaewon Jang, and S. Park. Daily activity recognition system for the elderly using pressure sensors. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 5188–5191, Aug 2008.

[25] Z. Uhrikova, C. D. Nugent, and V. Hlavac. The use of computer vision techniques to augment home based sensorised environments. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 2550–2553, Aug 2008.

[26] Sagar Shelke and Baris Aksanli. Static and dynamic activity detection with ambient sensors in smart spaces. Sensors, 19(4):804, 2019.

[27] Arindam Ghosh, Anubrata Sanyal, Amartya Chakraborty, Praveen Sharma, Mousumi Saha, Subrata Nandi, and Sujoy Saha. On automatizing recognition of multiple human activities using ultrasonic sensor grid. pages 488–491, 01 2017.

[28] Mathias Sundholm, Jingyuan Cheng, Bo Zhou, Akash Sethi, and Paul Lukowicz. Smart-Mat: Recognizing and Counting Gym Exercises with Low-cost Resistive Pressure Sensing Matrix. In A. J. Brush, Adrian Friday, Julie Kientz, James Scott, and Junehwa Song, editors, Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct, pages 373–382, New York, New York, USA, 2014. ACM Press.

[29] Rushil Khurana, Karan Ahuja, Zac Yu, Jennifer Mankoff, Chris Harrison, and Mayank Goel. Gymcam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2:1–17, 12 2018.

[30] Yang Qifan, Tang Hao, Zhao Xuebing, Li Yin, and Zhang Sanfeng. Dolphin: Ultrasonic-based gesture recognition on smartphone platform. In Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on, pages 1461–1468, Dec 2014.

[31] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingerio : Using sonar for fine-grained finger tracking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '16, page to appear. ACM, 2016.

[32] Robert Prance, Sam Beardsmore-Rust, P. Watson, C.J. Harland, and H. Prance. Remote detection of human electrophysiological signals using electric potential sensors. Applied Physics Letters, 93:033906–033906, 07 2008.

[33] Gabe Cohn, Daniel Morris, Shwetak Patel, and Desney Tan. Humantenna: Using the body as an antenna for real-time whole-body interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, page 1901–1910, New York, NY, USA, 2012. Association for Computing Machinery.

[34] Prashant Srinivasan, David Birchfield, Gang Qian, and Assegid Kidané. A pressure sensing floor for interactive media applications. In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACE '05, page 278–281, New York, NY, USA, 2005. Association for Computing Machinery.

[35] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: Ubiquitous gesture sensing with millimeter wave radar. ACM Transactions on Graphics (TOG), 35(4):142, 2016.

[36] Saiwen Wang, Jie Song, Jaime Lien, Ivan Poupyrev, and Otmar Hilliges. Interacting with soli: Exploring fine-grained dynamic gesture recognition in the radio-frequency spectrum. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16, page 851–860, New York, NY, USA, 2016. Association for Computing Machinery.

[37] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In 2015 IEEE Conference on Computer Communications (INFOCOM), pages 1472–1480. IEEE, 2015.

[38] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '15, page 45–57, New York, NY, USA, 2015. Association for Computing Machinery.

[39] Ashikur Rahman, Victor M Lubecke, Olga Boric-Lubecke, Jan H Prins, and Takuya Sakamoto. Doppler radar techniques for accurate respiration characterization and subject identification. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 8(2):350–359, 2018.

[40] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang. Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(3):121, 2019.

[41] Fadel Adib and Dina Katabi. See through walls with wifi! In Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM, SIGCOMM '13, page 75–86, New York, NY, USA, 2013. Association for Computing Machinery.

[42] Satoshi Nakamura, Kazuo Hiyane, Futoshi Asano, Takanobu Nishiura, and Takeshi Yamada. Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition. In LREC, 2000.

[43] Christophe Veaux, Junichi Yamagishi, and Simon King. The voice bank corpus: Design, collection and data analysis of a large regional accent speech database. pages 1–4, 11 2013.

[44] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. Indoor person identification through footstep induced structural vibration. In Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, HotMobile '15, pages 81–86, New York, NY, USA, 2015. ACM.

[45] Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. Characterizing wave propagation to improve indoor step-level person localization using floor vibration. In Jerome P. Lynch, editor, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2016, volume 9803, pages 30 – 40. International Society for Optics and Photonics, SPIE, 2016.

[46] Arindam Ghosh, Amartya Chakraborty, Dhruv Chakraborty, Mousumi Saha, and Sujoy Saha. Ultrasense: A non-intrusive approach for human activity identification using heterogeneous ultrasonic sensor grid for smart home environment. Journal of Ambient Intelligence and Humanized Computing, pages 1–22, 03 2019.

[47] Arindam Ghosh, Dhruv Chakraborty, Deepak Prasad, Mousumi Saha, and Sujoy Saha. Can we recognize multiple human group activities using ultrasonic sensors? pages 557–560, 01 2018.

[48] A. Patel, C. Prabhudesai, and B. Aksanli. Non-intrusive activity detection and prediction in smart residential spaces. In 2018 9th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON), pages 355–361, Nov 2018.

[49] Joshua R. Smith. Field mice: Extracting hand geometry from electric field measurements. IBM Systems Journal, 35:587–608, 1996.

[50] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. Touché: Enhancing touch interaction on humans, screens, liquids, and everyday objects. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, page 483–492, New York, NY, USA, 2012. Association for Computing Machinery.

[51] Andreas Braun, Sebastian Frank, Martin Majewski, and Xiaofeng Wang. Capseat: Capacitive proximity sensing for automotive activity recognition. In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '15, page 225–232, New York, NY, USA, 2015. Association for Computing Machinery.

[52] H. J. Lee, S. H. Hwang, S. M. Lee, Y. G. Lim, and K. S. Park. Estimation of body postures on bed using unconstrained ecg measurements. IEEE Journal of Biomedical and Health Informatics, 17(6):985–993, Nov 2013.

[53] Silvia Rus, Tobias Grosse-Puppendahl, and Arjan Kuijper. Recognition of bed postures using mutual capacitance sensing. In Emile Aarts, Boris de Ruyter, Panos Markopoulos, Evert van Loenen, Reiner Wichert, Ben Schouten, Jacques Terken, Rob Van Kranenburg, Elke Den Ouden, and Gregory O'Hare, editors, Ambient Intelligence, pages 51–66, Cham, 2014. Springer International Publishing.

[54] L. Faggion and A. E. Mahdi. Noncontact human electrophysiological measurements using a new displacement current sensor. In SENSORS, 2011 IEEE, pages 296–299, Oct 2011.

[55] Jun Rekimoto and Hua Wang. Sensing gamepad: electrostatic potential sensing for enhancing entertainment oriented interactions. pages 1457–1460, 01 2004.

[56] A. Pouryazdan, R. J. Prance, H. Prance, and D. Roggen. Wearable electric potential sensing: A new modality sensing hair touch and restless leg movement. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp '16, page 846–850, New York, NY, USA, 2016. Association for Computing Machinery.

[57] Denys Matthies, Bernhard Strecker, and Bodo Urban. Earfieldsensing: A novel in-ear electric field sensing to enrich wearable gesture input through facial expressions. 05 2017.

[58] Gabe Cohn, Sidhant Gupta, Tien-Jui Lee, Dan Morris, Joshua R. Smith, Matthew S. Reynolds, Desney S. Tan, and Shwetak N. Patel. An ultra-low-power human body motion sensor using static electric field sensing. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp '12, pages 99–102, New York, NY, USA, 2012. ACM.

[59] Timothy J Orsley. Capacitive touchscreen or touchpad for finger and active stylus, October 2 2012. US Patent 8,278,571.

[60] Lisa Scherf, Florian Kirchbuchner, Julian von Wilmsdorff, Biying Fu, Andreas Braun, and Arjan Kuijper. Step by step: Early detection of diseases using an intelligent floor. In European Conference on Ambient Intelligence, pages 131–146. Springer, 2018.

[61] Samiksha Sikarwar, Satyendra, Shakti Singh, and Bal Chandra Yadav. Review on pressure sensors for structural health monitoring. Photonic Sensors, 7(4):294–304, 2017.

[62] Wenyao Xu, Ming-Chun Huang, Navid Amini, Lei He, and Majid Sarrafzadeh. ecushion: A textile pressure sensor array design and calibration for sitting posture analysis. IEEE Sensors Journal, 13(10):3926–3934, 2013.

[63] M. Rofouei, W. Xu, and M. Sarrafzadeh. Computing with uncertainty in a smart textile surface for object recognition. In 2010 IEEE Conference on Multisensor Fusion and Integration, pages 174–179, Sep. 2010.

[64] M. D. Addlesee, A. Jones, F. Livesey, and F. Samaria. The orl active floor [sensor system]. IEEE Personal Communications, 4(5):35–41, Oct 1997.

[65] G. Feng, J. Mai, Z. Ban, X. Guo, and G. Wang. Floor pressure imaging for fall detection with fiber-optic sensors. IEEE Pervasive Computing, 15(2):40–47, Apr 2016.

[66] Michiko Nishiyama, Hiroyuki Sasaki, Shinichi Nose, Kazumasa Takami, and Kazuhiro Watanabe. Smart pressure sensing mats with embedded hetero-core fiber optic nerve sensors. Materials and Manufacturing Processes, 25(4):264–267, 2010.

[67] L. Wang, S. Gupta, K. J. Loh, and H. S. Koo. Distributed pressure sensing using carbon nanotube fabrics. IEEE Sensors Journal, 16(12):4663–4664, June 2016.

[68] Gang Qian, Jiqing Zhang, and Assegid Kidané. People identification using gait via floor pressure sensing and analysis. In Daniel Roggen, Clemens Lombriser, Gerhard Tröster, Gerd Kortuem, and Paul Havinga, editors, Smart Sensing and Context, pages 83–98, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[69] Selene Mota and Rosalind W Picard. Automated posture analysis for detecting learner's interest level. In 2003 Conference on Computer Vision and Pattern Recognition Workshop, volume 5, pages 49–49. IEEE, 2003.

[70] Ahmed M. Almassri, W. Z. Wan Hasan, S. A. Ahmad, A. J. Ishak, A. M. Ghazali, D. N. Talib, and Chikamune Wada. Pressure Sensor: State of the Art, Design, and Application for Robotic Hand. Journal of Sensors, 2015(1):1–12, 2015.

[71] Robert Orr and Gregory Abowd. The smart floor: A mechanism for natural user identification and tracking. 03 2000.

[72] Sanjoy Dasgupta. Learning mixtures of gaussians. In 40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039), pages 634–644. IEEE, 1999.

[73] Shobhit Saxena, François Brémond, Monnique Thonnat, and Ruihua Ma. Crowd behavior recognition for video surveillance. In Jacques Blanc-Talon, Salah Bourennane, Wilfried Philips, Dan Popescu, and Paul Scheunders, editors, Advanced Concepts for Intelligent Vision Systems, pages 970–981, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[74] K. Martin Sagayam and D. Jude Hemanth. Hand posture and gesture recognition techniques for virtual reality applications: a survey. Virtual Reality, 21(2):91–107, Jun 2017.

[75] Wei-Lwun Lu and James J. Little. Simultaneous tracking and action recognition using the pca-hog descriptor. In Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision, CRV '06, page 6, USA, 2006. IEEE Computer Society.

[76] Ying Luo, Tzong-Der Wu, and Jenq-Neng Hwang. Object-based analysis and interpretation of human motion in sports video sequences by dynamic bayesian networks. Computer Vision and Image Understanding, 92:196–216, 11 2003.

[77] Xin Lu, Qiong Liu, and Shunichiro Oe. Recognizing non-rigid human actions using joints tracking in space-time. In Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04) Volume 2 - Volume 2, ITCC '04, page 620, USA, 2004. IEEE Computer Society.

[78] Baris Kayalibay, Grady Jensen, and Patrick van der Smagt. Cnn-based segmentation of medical imaging data. arXiv preprint arXiv:1701.03056, 2017.

[79] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. Mask R-CNN. CoRR, abs/1703.06870, 2017.

[80] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In arXiv preprint arXiv:1812.08008, 2018.

[81] A. Ali and J. K. Aggarwal. Segmentation and recognition of continuous human activity. In Proceedings IEEE Workshop on Detection and Recognition of Events in Video, pages 28–35, July 2001.

[82] Debapratim Das Dawn and Soharab Hossain Shaikh. A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector. Vis. Comput., 32(3):289–306, March 2016.

[83] T. V. Nguyen, Z. Song, and S. Yan. Stap: Spatial-temporal attention-aware pooling for action recognition. IEEE Transactions on Circuits and Systems for Video Technology, 25(1):77–86, Jan 2015.

[84] Limin Wang, Yu Qiao, and Xiaoou Tang. Action recognition with trajectory-pooled deep-convolutional descriptors. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[85] Guangyu Zhu, Qingming Huang, Changsheng Xu, Yong Rui, Shuqiang Jiang, Wen Gao, and Hongxun Yao. Trajectory based event tactics analysis in broadcast sports video. In Proceedings of the 15th ACM International Conference on Multimedia, MM '07, page 58–67, New York, NY, USA, 2007. Association for Computing Machinery.

[86] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12, pages 1097–1105, USA, 2012. Curran Associates Inc.

[87] Andrew Ng et al. Sparse autoencoder. CS294A Lecture notes, 72(2011):1–19, 2011.

[88] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks, 2015. cite arxiv:1511.06434Comment: Under review as a conference paper at ICLR 2016.

[89] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), ICCV '15, page 4489–4497, USA, 2015. IEEE Computer Society.

[90] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-scale video classification with convolutional neural networks. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, pages 1725–1732, June 2014.

[91] Khurram Soomro, Amir Roshan Zamir, and Mubarak Shah. UCF101: A dataset of 101 human actions classes from videos in the wild. CoRR, abs/1212.0402, 2012.

[92] K. G. Derpanis, M. Lecce, K. Daniilidis, and R. P. Wildes. Dynamic scene understanding: The role of orientation features in space and time in scene classification. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 1306–1313, June 2012.

[93] Ali Diba, Ali Mohammad Pazandeh, and Luc Van Gool. Efficient two-stream motion and appearance 3d cnns for video classification. CoRR, abs/1608.08851, 2016.

[94] Matthias Ruben Mettel, Michael Alekseew, Carsten Stocklöw, and Andreas Braun. Designing and evaluating safety services using depth cameras. Journal of Ambient Intelligence and Humanized Computing, 10(2):747–759, 2019.

[95] Enea Cippitelli, Samuele Gasparrini, Ennio Gambi, and Susanna Spinsante. A human activity recognition system using skeleton data from rgbd sensors. Computational intelligence and neuroscience, 2016, 2016.

[96] Salvatore Gaglio, Giuseppe Lo Re, and Marco Morana. Human activity recognition process using 3-d posture data. IEEE Transactions on Human-Machine Systems, 45(5):586–597, 2014.

[97] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena. Unstructured human activity detection from rgbd images. In 2012 IEEE international conference on robotics and automation, pages 842–849. IEEE, 2012.

[98] Lu Xia, Chia-Chih Chen, and Jake K Aggarwal. View invariant human action recognition using histograms of 3d joints. In 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pages 20–27. IEEE, 2012.

[99] Lorenzo Seidenari, Vincenzo Varano, Stefano Berretti, Alberto Bimbo, and Pietro Pala. Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 479–485, 2013.

[100] Wanqing Li, Zhengyou Zhang, and Zicheng Liu. Action recognition based on a bag of 3d points. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pages 9–14. IEEE, 2010.

[101] Ju Han and Bir Bhanu. Human activity recognition in thermal infrared imagery. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, pages 17–17. IEEE, 2005.

[102] Adam Williams, Deepak Ganesan, and Allen Hanson. Aging in place: fall detection and localization in a distributed smart camera network. pages 892–901, 01 2007.

[103] S. L. Dockstade and A. M. Tekalp. Multiple camera tracking of interacting and occluded human motion. Proceedings of the IEEE, 89(10):1441–1455, Oct 2001.

[104] Jay Prakash Gupta, Nishant Singh, Pushkar Dixit, Vijay Bhaskar Semwal, and Shiv Ram Dubey. Human activity recognition using gait pattern. International Journal of Computer Vision and Image Processing (IJCVIP), 3(3):31–53, 2013.

[105] Luigi Di Stefano and Andrea Bulgarelli. A simple and efficient connected components labeling algorithm. In Proceedings 10th International Conference on Image Analysis and Processing, pages 322–327. IEEE, 1999.

[106] Peter Hevesi, Sebastian Wille, Gerald Pirkl, Norbert Wehn, and Paul Lukowicz. Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, pages 141–145, 2014.

[107] Takayuki Kawashima, Yasutomo Kawanishi, Ichiro Ide, Hiroshi Murase, Daisuke Deguchi, Tomoyoshi Aizawa, and Masato Kawade. Action recognition from extremely low-resolution thermal image sequence. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pages 1–6. IEEE, 2017.

[108] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.

[109] Mohammad Javad Shafiee, Brendan Chywl, Francis Li, and Alexander Wong. Fast yolo: A fast you only look once system for real-time embedded object detection in video. arXiv preprint arXiv:1709.05943, 2017.

[110] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015.

[111] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.

[112] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 2961–2969, 2017.

[113] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. Eye-mms: Miniature multi-scale segmentation network of key eye-regions in embedded applications. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 3665–3670, Oct 2019.

[114] Pedro OO Pinheiro, Ronan Collobert, and Piotr Dollár. Learning to segment object candidates. In Advances in Neural Information Processing Systems, pages 1990–1998, 2015.

[115] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3156–3164, 2017.

[116] Yong Du, Wei Wang, and Liang Wang. Hierarchical recurrent neural network for skeleton based action recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1110–1118, 2015.

[117] Naser Damer, Fadi Boutros, Khawla Mallat, Florian Kirchbuchner, Jean-Luc Dugelay, and Arjan Kuijper. Cascaded generation of high-quality color visible face images from thermal captures. CoRR, abs/1910.09524, 2019.

[118] Asit Mishra and Debbie Marr. Apprentice: Using knowledge distillation techniques to improve low-precision network accuracy. arXiv preprint arXiv:1711.05852, 2017.

[119] Babak Hassibi and David G Stork. Second order derivatives for network pruning: Optimal brain surgeon. In Advances in neural information processing systems, pages 164–171, 1993.

[120] Marc Langheinrich. A Privacy Awareness System for Ubiquitous Computing Environments. In Gaetano Borriello and Lars Erik Holmquist, editors, UbiComp 2002: Ubiquitous Computing, pages 237–245, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[121] H. Rohling, S. Heuel, and H. Ritter. Pedestrian detection procedure integrated into an 24 ghz automotive radar. In 2010 IEEE Radar Conference, pages 1229–1232, May 2010.

[122] F. Folster, H. Rohling, and U. Lubbert. An automotive radar network based on 77 ghz fmcw sensors. In IEEE International Radar Conference, 2005., pages 871–876, May 2005.

[123] Dionysios Skordoulis, Qiang Ni, Hsiao-Hwa Chen, Adrian P Stephens, Changwen Liu, and Abbas Jamalipour. Ieee 802.11 n mac frame aggregation mechanisms for next-generation high-throughput wlans. IEEE Wireless Communications, 15(1):40–47, 2008.

[124] Richard van Nee and Ramjee Prasad. OFDM for wireless multimedia communications. Artech House, Inc., 2000.

[125] Mohan Liyanage, Chii Chang, Satish Srirama, and Seng Loke. Indoor people density sensing using wi-fi and channel state information. Advances in modelling and analysis A, 61(1):37–47, 2018.

[126] Bohn Dieter. Google's project soli: The tech behind pixel 4's motion sense radar. The VERGE.

[127] Ann-Kathrin Seifert, Moeness G Amin, and Abdelhak M Zoubir. Toward unobtrusive in-home gait analysis based on radar micro-doppler signatures. IEEE Transactions on Biomedical Engineering, 66(9):2629–2640, 2019.

[128] Liang Liu, Mihail Popescu, Marilyn Rantz, and Marjorie Skubic. Fall detection using doppler radar and classifier fusion. In Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, pages 180–183. IEEE, 2012.

[129] Beth Logan et al. Mel frequency cepstral coefficients for music modeling. In Ismir, volume 270, pages 1–11, 2000.

[130] Youngwook Kim and Taesup Moon. Human detection and activity classification based on micro-doppler signatures using deep convolutional neural networks. IEEE geoscience and remote sensing letters, 13(1):8–12, 2015.

[131] Armin Schneider and Hubertus Feussner. Chapter 11 - tracking and navigation systems. In Armin Schneider and Hubertus Feussner, editors, Biomedical Engineering in Gastrointestinal Surgery, pages 443 – 472. Academic Press, 2017.

[132] Chieh-Ping Lai and Ram M Narayanan. Through-wall imaging and characterization of human activity using ultrawideband (uwb) random noise radar. In Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV, volume 5778, pages 186–195. International Society for Optics and Photonics, 2005.

[133] Chuanwei Ding, Li Zhang, Chen Gu, Lei Bai, Zhicheng Liao, Hong Hong, Yusheng Li, and Xiaohua Zhu. Non-contact human motion recognition based on uwb radar. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 8(2):306–315, 2018.

[134] Wei Wang, Alex X Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pages 363–373. ACM, 2016.

[135] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In Proceedings of the 19th Annual International Conference on Mobile Computing & Networking, MobiCom '13, page 27–38, New York, NY, USA, 2013. Association for Computing Machinery.

[136] F Zafari, A Gkelias, and K Leung. A survey of indoor localization systems and technologies. arxiv 2017. arXiv preprint arXiv:1709.01015.

[137] Rong Liu and Ming Liu. Recognizing human activities based on multi-sensors fusion. In 2010 4th International Conference on Bioinformatics and Biomedical Engineering, pages 1–4. IEEE, 2010.

[138] Chun Zhu and Weihua Sheng. Human daily activity recognition in robot-assisted living using multi-sensor fusion. In 2009 IEEE International Conference on Robotics and Automation, pages 2154–2159. IEEE, 2009.

[139] Moritz Tenorth, Jan Bandouch, and Michael Beetz. The tum kitchen data set of everyday manipulation activities for motion tracking and action recognition. In 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, pages 1089–1096. IEEE, 2009.

[140] Rasika S Ransing and Manita Rajput. Smart home for elderly care, based on wireless sensor network. In 2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE), pages 1–5. IEEE, 2015.

[141] Stanislav Safaric and Kresimir Malaric. Zigbee wireless standard. In Proceedings ELMAR 2006, pages 259–262. IEEE, 2006.

[142] Kerem Altun and Billur Barshan. Human activity recognition using inertial/magnetic sensor units. In International workshop on human behavior understanding, pages 38–51. Springer, 2010.

[143] U. Maurer, A. Smailagic, D.P. Siewiorek, and Michael Deisher. Activity recognition and monitoring using multiple sensors on different body positions. volume 2006, pages 4 pp.–, 05 2006.

[144] Ling Bao and Stephen S. Intille. Activity recognition from user-annotated acceleration data. In Alois Ferscha and Friedemann Mattern, editors, Pervasive, volume 3001 of Lecture Notes in Computer Science, pages 1–17. Springer, 2004.

[145] Saisakul Chernbumroong, Shuang Cang, and Hongnian Yu. Genetic algorithm-based classifiers fusion for multisensor activity recognition of elderly people. IEEE journal of biomedical and health informatics, 19(1):282–289, 2014.

[146] David E Goldberg and John Henry Holland. Genetic algorithms and machine learning. 1988.

[147] Aniesha Alford, Caresse Hansen, Gerry V. Dozier, Kelvin S. Bryant, John C. Kelly, Tamirat Abegaz, Karl Ricanek, and Damon L. Woodard. Gec-based multi-biometric fusion. In Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2011, New Orleans, LA, USA, 5-8 June, 2011, pages 2071–2074. IEEE, 2011.

[148] Naser Damer, Chadi Izzou Rhaibani, Andreas Braun, and Arjan Kuijper. Trust the biometrie mainstream: Multi-biometric fusion and score coherence. In 25th European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, August 28 - September 2, 2017, pages 2191–2195. IEEE, 2017.

[149] Naser Damer, Alexander Opel, and Alexander Nouak. Biometric source weighting in multi-biometric fusion: Towards a generalized and robust solution. In 22nd European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal, September 1-5, 2014, pages 1382–1386. IEEE, 2014.

[150] Naser Damer and Alexander Opel. Multi-biometric score-level fusion and the integration of the neighbors distance ratio. In Aurélio J. C. Campilho and Mohamed S. Kamel, editors, Image Analysis and Recognition - 11th International Conference, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part II, volume 8815 of Lecture Notes in Computer Science, pages 85–93. Springer, 2014.

[151] Naser Damer. Application-driven Advances in Multi-biometric Fusion. PhD thesis, Darmstadt University of Technology, Germany, 2018.

[152] D. Bellot, A. Boyer, and F. Charpillet. A new definition of qualified gain in a data fusion process: application to telemedicine. In Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997), volume 2, pages 865–872 vol.2, July 2002.

[153] Daniel Roggen, Alberto Calatroni, Mirco Rossi, Thomas Holleczek, Kilian Forster, Gerhard Troster, Paul Lukowicz, David Bannach, Gerald Pirkl, Alois Ferscha, Jakob Doppler, Clemens Holzmann, Marc Kurz, Gerald Holl, Ricardo Chavarriaga, Hesam Sagha, Hamidreza Bayati, Marco Creatura, and Jose del R. Millan. Collecting complex activity datasets in highly rich networked sensor environments. pages 233 – 240, 07 2010.

[154] Billur Barshan and Murat Yuksek. Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. The Computer Journal, 57:1649–1667, 10 2013.

[155] A. Reiss and D. Stricker. Introducing a new benchmarked dataset for activity monitoring. In 2012 16th International Symposium on Wearable Computers, pages 108–109, June 2012.

[156] Tim van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. Accurate activity recognition in a home setting. In Proceedings of the 10th International Conference on Ubiquitous Computing, UbiComp '08, pages 1–9, New York, NY, USA, 2008. ACM.

[157] Stephen Intille, Kent Larson, Emmanuel Tapia, Jennifer Beaudin, Pallavi Kaushik, Jason Nawyn, and Randy Rockinson. Using a live-in laboratory for ubiquitous computing research. volume 3968, pages 349–365, 05 2006.

[158] Donald Patterson, Dieter Fox, Henry Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. volume 2005, pages 44 – 51, 11 2005.

[159] Meinard Müller, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. Documentation mocap database hdm05. 2007.

[160] Christian Schuldt, Ivan Laptev, and Barbara Caputo. Recognizing human actions: a local svm approach. In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., volume 3, pages 32–36. IEEE, 2004.

[161] Jiang Wang, Zicheng Liu, Ying Wu, and Junsong Yuan. Mining actionlet ensemble for action recognition with depth cameras. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 1290–1297. IEEE, 2012.

[162] Ross Messing, Chris Pal, and Henry Kautz. Activity recognition using the velocity histories of tracked keypoints. In 2009 IEEE 12th international conference on computer vision, pages 104–111. IEEE, 2009.

[163] Fernando De la Torre, Jessica Hodgins, Javier Montano, Sergio Valcarcel, R Forcada, and J Macey. Guide to the carnegie mellon university multimodal activity (cmu-mmac) database. Robotics Institute, Carnegie Mellon University, 5, 2009.

[164] Marcus Rohrbach, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele. A database for fine grained activity detection of cooking activities. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 1194–1201. IEEE, 2012.

[165] Pradeep Natarajan and Ramakant Nevatia. View and scale invariant action recognition using multiview shape-flow models. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2008.

[166] Ronald Poppe. A survey on vision-based human action recognition. Image and vision computing, 28(6):976–990, 2010.

[167] Pavan Turaga, Rama Chellappa, Venkatramana S Subrahmanian, and Octavian Udrea. Machine recognition of human activities: A survey. IEEE Transactions on Circuits and Systems for Video technology, 18(11):1473–1488, 2008.

[168] Long Cheng, Yiyi Yu, Xinyang Liu, Jinyu Su, and Yani Guan. Recognition of human activities using fast and adaptive sparse representation based on wearable sensors. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 944–949. IEEE, 2017.

[169] Long Cheng, Yiyang Li, and Yani Guan. Human activity recognition based on compressed sensing. In 2017 IEEE 7th annual computing and communication workshop and conference (CCWC), pages 1–7. IEEE, 2017.

[170] Aiguo Wang, Guilin Chen, Cuijuan Shang, Miaofei Zhang, and Li Liu. Human activity recognition in a smart home environment with stacked denoising autoencoders. In International conference on web-age information management, pages 29–40. Springer, 2016.

[171] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on multi-channel time series for human activity recognition. In Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[172] Thomas Plötz, Nils Y Hammerla, and Patrick L Olivier. Feature learning for activity recognition in ubiquitous computing. In Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

[173] Lourdes Martínez-Villaseñor, Hiram Ponce, and Ricardo Abel Espinosa-Loera. Multimodal database for human activity recognition and fall detection. In Multidisciplinary Digital Publishing Institute Proceedings, volume 2, page 1237, 2018.

[174] Lasitha Piyathilaka and Sarath Kodagoda. Gaussian mixture based hmm for human daily activity recognition using 3d skeleton features. In 2013 IEEE 8th conference on industrial electronics and applications (ICIEA), pages 567–572. IEEE, 2013.

[175] Ehsan Nazerfard, Barnan Das, Lawrence B Holder, and Diane J Cook. Conditional random fields for activity recognition in smart environments. In Proceedings of the 1st ACM International Health Informatics Symposium, pages 282–286. ACM, 2010.

[176] Mark V Albert, Santiago Toledo, Mark Shapiro, and Konrad Koerding. Using mobile phones for activity recognition in parkinson's patients. Frontiers in neurology, 3:158, 2012.

[177] Homay Danaei Mehr, Huseyin Polat, and Aydin Cetin. Resident activity recognition in smart homes by using artificial neural networks. In 2016 4th international istanbul smart grid congress and fair (ICSG), pages 1–5. IEEE, 2016.

[178] Maja Stikic, Kristof Van Laerhoven, and Bernt Schiele. Exploring semi-supervised and active learning for activity recognition. In 2008 12th IEEE International Symposium on Wearable Computers, pages 81–88. IEEE, 2008.

[179] Alexandros Iosifidis, Ermioni Marami, Anastasios Tefas, and Ioannis Pitas. Eating and drinking activity recognition based on discriminant analysis of fuzzy distances and activity volumes. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 2201–2204. IEEE, 2012.

[180] Jamie A Ward, Paul Lukowicz, and Hans W Gellersen. Performance metrics for activity recognition. ACM Transactions on Intelligent Systems and Technology (TIST), 2(1):1–23, 2011.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.2991891, IEEE Access

[181] Yan Ke, Rahul Sukthankar, and Martial Hebert. Spatio-temporal shape and flow correlation for action recognition. In 2007 IEEE conference on computer vision and pattern recognition, pages 1–8. IEEE, 2007.

[182] Johannes A Stork, Luciano Spinello, Jens Silva, and Kai O Arras. Audio-based human activity recognition using non-markovian ensemble voting. In 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, pages 509–514. IEEE, 2012.

[183] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '11, page 1297–1304, USA, 2011. IEEE Computer Society.

BIYING FU received the B.S. degree in electrical engineering and information technoglogy from University of Karlsruhe, Germany, in 2011 and the M.S. degree in electrical engineering and information technoglogy from University of Karlsruhe, in 2014. She is currently pursuing the Ph.D. degree in informatics at Technical University of Darmstadt, Germany, on the topic "Applications for Human Activity Recognition with Non-optical Sensors". From 2014 till today, she is working as a Research Assistant with the Fraunhofer Institut for Computer Graphics IGD, Germany. Her research interest includes intelligent human machine interaction, machine learning, and deep learning for human activity recognition with sensor data. She is currently working in the department for Smart Living & Biometric Technologies.

NASER DAMER is a senior researcher at the competence center Smart Living & Biometric Technologies, Fraunhofer IGD. He received his master of science degree in electrical engineering from the Technical University of Kaiserslautern (2010) and his PhD in computer science from the Technical University of Darmstadt (2018). He is a researcher at Fraunhofer IGD since 2011 performing applied research, scientific consulting, and system evaluation. His main research interests lie in the fields of biometrics, machine learning and information fusion. He published more than 60 scientific papers in these fields. Dr. Damer is a Principal Investigator at the National Research Center for Applied Cybersecurity ATHENE in Darmstadt, Germany. He serves as a reviewer for a number of journals and conferences and as an associate editor for the Visual Computer journal. He represents the German Institute for Standardization (DIN) in ISO/IEC SC37 biometrics standardization committee.

FLORIAN KIRCHBUCHNER is trained as an information and telecommunication systems technician and served as IT expert for the German Army from 2001 to 2009. Afterwards he studied computer science at Technical University of Darmstadt and graduated with a Master of Science degree in 2014. He has been working at Fraunhofer IGD since 2014, most recently as head of the department for Smart Living & Biometric Technologies. He is also Principal Investigator at the National Research Center for Applied Cybersecurity ATHENE. Mr. Kirchbuchner participated at Software Campus, a management program of the Federal Ministry of Education and Research (BMBF) and is currently doing his PhD at Technical University of Darmstadt on the topic "Electric Field Sensing for Smart Support Systems: Applications and Implications".

ARJAN KUIJPER received the M.Sc. degree in applied mathematics from Twente University, The Netherlands, the Ph.D. degree from UtrechtUniversity, The Netherlands, and the Habitation degree from TU Graz, Austria. He was an Assistant Research Professor with the IT University of Copenhagen, Denmark, and a Senior Researcher with RICAM, Linz, Austria. His research interests include all aspects of mathematics-based methods for computer vision, graphics, imaging, pattern recognition, interaction, and visualization. He is a member of the management of Fraunhofer IGD, where he is responsible for scientific dissemination. He holds the Chair in mathematical and applied visual computing with TU Darmstadt. He is the author of over 300 peer-reviewed publications,the Associate Editor of CVIU, PR, and TVCJ, the Secretary of the International Association for Pattern Recognition (IAPR), and serves both as a Reviewer for many journals and conferences, and as a program committee member and organizer of conferences.