Chapter 12
Contactless Human Monitoring: Challenges and Future Direction

Upal Mahbub, Tauhidur Rahman, and Md Atiqur Rahman Ahad

Abstract  Human activity recognition and analysis have a great number of important applications in numerous fields including computer vision, ubiquitous computing, human-computer interactions, healthcare, robotics, and surveillance. Video-based and sensor-based human activity recognition have progressed tremendously in the last two decades. In this chapter, we highlight the major challenges and future research directions in contactless human activity monitoring. More specifically, we will present sensor-level challenges, feature-level challenges, methodological issues, implementation-level aspects, and various application-level challenges.

Keywords  Human activity recognition · Contactless sensing · Activity recognition challenges · Future of activity monitoring

12.1 Introduction

Monitoring human activities in a contact-free manner is a holy-grail in ubiquitous computing that can unlock numerous applications. Researchers in the areas of Computer Science, Electrical Engineering and Informatics have been searching for novel, unobtrusive, privacy-preserving and high-fidelity approaches for passively monitoring everyday activities that can be implemented in resource constrained and noisy environment. The fundamental gap in contactless human activity recognition is highlighted by the fact that existing approaches largely fail to provide us scalable solu-
tions that can achieve high performance, while being unobtrusive. Moreover, several confounding factors including the wide variations in environment, presence of occlusion, variation in viewpoint and appearances, large intra-class differences, fuzzy class boundaries, unforeseen events, the presence of motion and noise further complicates the activity analysis [1–4].

Recent studies [5, 6] have suggested that it is critical to take an interdisciplinary approach for designing an automated human activity recognition (HAR) system. In [7], the authors classified the task of HAR into three parts:

- Human activity detection and localization,
- Human activity modeling, i.e., feature extraction, and
- Human activity classification.

While this classification highlights the modeling and machine learning related aspects, an End-to-End contactless human activity recognition system requires careful considerations on real-world hardware implementation, efficient and resource constraint computation, novel physics-based sensing, socio-economical and human factor related challenges.

Contactless Human Activity Analysis (CHAA) aims to go beyond a general activity recognition system with the additional contactlessness/off-body requirement. Building off-body contactless sensing technology is the core highlight of this area which typically makes the problem of activity recognition significantly more challenging. The technical challenges for the design and development of contactless sensors include designing new sensors or sensing approaches, designing an effective data collection scheme, the efficiency and accuracy of the algorithm, hardware implementation for real-time operation, and privacy considerations.

In this chapter, we systematically divide the CHAA-related challenges into five levels, which are—

1. Sensor-level challenges
2. Feature-level challenges
3. Algorithm-level challenges
4. Implementation-level challenges, and,
5. Application-level challenges.

In Fig. 12.1, we provide the lists of topics that we will cover in this chapter. We endeavor to discuss the challenges and future research directions from different angles to give a comprehensive idea about the broad spectrum of research possibilities in this field to the readers. In Sects. 12.2–12.6, we will shed light on each level, respectively, to highlight the relevant challenges and future research directions for that particular level.
12.2 Sensor-Level Challenges

A wide range of sensors, starting from video/image sensors (RGB, depth maps) to numerous wearable and environmental sensors, have been proposed and used for HAR tasks. Since, the scope of this chapter is contactless human activity analysis, the challenges related to wearable sensor-based activity recognition are not discussed here. The core challenges related to sensor-related schemes are how to capture real-world data accurately, how to define the optimum number and categories of sensors, how to encompass large-scale data for research and analysis, ensuring privacy of the data, placement of the sensors for smarter analysis, handling multiple sensors, and computing capacity issues in terms of latency, power and bandwidth.

12.2.1 Accurate Sensing of the Real-World

When capturing the real-world using sensors, there are always artifacts present in the data in the form of noise. Noise can come from the sensor itself because of motion, quantization, or data capturing mechanism. The environment can also be noisy; for example, speech data collected in a factory may have heavy noise in the
background. For image or video data, there can be noise in the image due to rain, snow, or low-visibility [8], over/under exposure depending on lighting condition, distortion because of the medium such as underwater photographs [9] or camera type (e.g., large field of view cameras with lens distortion [10]) [11]. Artifacts can occur due to the motion of the subject in the scene or the sensor itself. Noise-free capture of real-world information is nearly impossible. Therefore, the challenge is to process the sensor data in a meaningful way to reduce/remove noise and compensate for the distortion. The impact of noise in recognition tasks is known to be significant, as showed in many studies over the years [12, 13].

In recent years, sensor data collection has tilted more towards multi-modal systems. Research on noise reduction and source separation using sensor arrays (e.g., microphone arrays used in Amazon Echo devices), fusing different types of sensors such as radar, lidar, RGB plus depth cameras, and near-infrared to capture fine details of a scene (e.g., in self-driving cars), and using deep learning-based preprocessing for cleaning up or enhancing raw data are now gaining traction.

12.2.2 Optimum Number and Types of Sensors

Major works on human activity analysis focus on single-sensor, single-user, and single-activity. However, most real-world use cases involve multiple subjects doing multiple-activities and are best described using several modalities such as multi-view setups or different types of sensor data, such as video, audio, and depth. With the rapid advancement of computational power and sensor technologies, it is now apparent that a combination of multiple sensors is capable of capturing human activities more accurately and robustly in comparison to single-sensor systems. Yet, the optimum number and types of sensors required to capture an action entirely are still debatable.

If we consider humans as ideal activity recognition machines, then we can argue that a stereo-vision system similar to the eyes and two spatially separated audio sensors like the ears should be sufficient. However, in reality, not only human bodies can sense touch, smell, and vibration, but humans can also effectively use prior experiences and contexts in a versatile way to interpret and understand a scene. Moreover, the types of sensors to be used for analyzing an action is currently very task-specific. For example, most activity datasets are video-based (RGB, depth and/or skeleton data), where audio and text are rarely incorporated. Also, the sizes of the multi-modal datasets are usually smaller in terms of the number of classes, viewpoints, and scenarios. Thus, a challenging future work in this domain would be to provide a generic guideline for optimum number and type of sensors based on extensive research on large datasets collected in ‘in the wild.’
12.2.3 Large Scale Data Collection

In [7], the authors suggested that an ideal dataset for HAR should have a sufficient amount of images and/or video sequences as input media and address several issues, namely, the quality/resolution of the input media, a large variety of subjects, action classes, illumination, and large intra-class variance. Also, the background should be complex, and the dataset should have partially occluded human structures. However, how large is large enough for an action dataset is an open question—especially now that deep neural networks can leverage large datasets and most methods specifically rely on large labeled data [11, 14]. Also, as discussed in [11], collecting large action dataset is a cumbersome task, and even more laborious is to get the data annotated. Moreover, large scale annotated datasets are usually created while focusing on a specific task, whereas multi-modal approaches are proven to be more effective. For example, in [15], the authors showed that a multi-task learning framework helps to improve the performance of a deep neural network for all individual tasks. However, to design that multi-tasking framework, the authors had to train using six different medium-to-large datasets, each covering different data modalities. Some datasets have annotations for age and gender, others for identity and/or attribute like age, smile, some had annotations for head pose and detection bounding boxes [16].

Some concepts, like multi-modal active authentication of smartphones, rely on fusing results from different modalities such as camera capture, audio, inertial sensor information, application-usage, or location-sensor data [17–20]. Still, there is no publicly available dataset that is large enough to develop and evaluate such systems reliably. Most available datasets are task-specific and focused on a subset of the available modalities such as the Brain Run Behavioral Biometric Dataset [21] and H-MOG dataset [22]; on the other hand, the datasets are wide-ranging, realistic, but very small, such as the UMDAA-02 dataset [17]. Another interesting example is the Sports-1M dataset, a huge automatic unconstrained sports action recognition dataset [14]. The dataset clearly suffers from label noise because of the automatic collection protocol, has very low intra-class variation, partially unavailable because the original users removed the videos from youtube and the input media is only videos [23]. In a nutshell, there is clearly a lack of consolidated benchmark datasets for exhaustive training of deep learning models [24].

On the other hand, training a deep network with a small dataset is another major challenge [11]. There are many techniques to avoid overfitting and force the network to generalize better, such as custom regularization, dropout, transfer learning. Also, unsupervised learning with large-scale unlabeled data in conjunction with supervised training with a small amount of labeled data, popularly known as semi-supervised learning, helps reduce the over-reliance on labeled data for activity recognition using deep learning [25, 26]. Techniques to harness information from large unlabeled or weakly labeled data for contactless sensing applications is an open challenge.
12.2.4 Privacy Issues

Sensor data such as camera capture, an audio stream, or physiological information can contain sensitive information related to user’s privacy. It is imperative to ensure confidentiality in the human activity analysis process to secure a user’s privacy. In this regard, research on reliable data anonymization process to secure a user’s privacy is a big challenge. For example, in [27], the authors presented a method to anonymize the faces of doctors and patients when capturing videos in the surgical operating room. Privacy-preserving activity recognition can be a concern for smart homes with many cameras. In [28], the authors simulated using extremely low temporal and spatial resolution cameras to ensure privacy preserved interaction with human subjects. The results show that action recognition performance drops significantly if the environment is degraded to such an extreme setup. Considering the growing importance of privacy preserving data analysis, the research branch of Differential Privacy has evolved which, according to [29], ...describes a promise, made by a data holder or curator, to a data subject: “You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available.” In [29] the authors delved into the mathematical foundations of differential privacy and addressed key topics such as differentially private local and online machine learning. In this regard, an interesting challenge is to quantify the additional sample and run time required to privately learn, as compared with the same for learning non-privately.

12.2.5 Sensor Placement Challenges

Egocentric-view has important aspects in some applications. First-person point-of-view cameras can provide different kinds of data, e.g., finger movements, hand movements when the camera is mounted on the forehead, eye-level, neck-level, or chest. For example, automatic hand detection, hand pose estimation, and tracking in most likely going to be an integral part of all virtual reality and augmented reality applications in the future. Hand tracking is recently introduced as a feature in the Oculus Quest VR device [30]. Rehabilitation, law enforcement, extreme sports, etc. are a few related applications based on egocentric-view-based activity understanding. Note that even though the cameras are attached to the body or clothes, these special use cases can still be considered as contactless depending on the problem setup.

Apart from that, we can setup multiple cameras for video surveillance or recording. Smart homes and elderly support systems are incorporated with various kind of sensors with varied locations. How to optimize the placement that can provide smarter less-occluded data while retaining less data storage issue—is another important challenging point. Moreover, for a few applications, we need both contact-based wearable sensors as well as contactless sensors—to work together. How to deal this issue is an important research point for the future. While the collaboration between
on-body and off-body sensor data has been explored, online algorithms that compensate for missing/noisy data in each sensor stream with different active learning schemes is still an open problem.

### 12.2.6 Resolution, Bandwidth and Power

Operating a sensor at high resolution would mostly cause lower bandwidth and high power consumption. Finding the proper balance between resolution, bandwidth, and power has always been a challenging task. An interesting direction might be combining multiple low-resolution sensor outputs to obtain better signal resolution while operating at a relatively high bandwidth and low power [31]. Also, introducing learning-based techniques as a post-processing step to improve the quality of the captured low-resolution signal is an interesting challenge in this domain. For example, in [32], the authors performed image super-resolution on the capture of a low-resolution imaging sensor for lens-less blood-cell counting.

Moreover, for some applications or modalities, we do not need continuous data storage or transfer or processing. Therefore, how to manage smartly is an issue. There are few ubiquitous sensor-based works where authors have tried to optimize the data collections. For example, if a person is not moving in a room, then a video data can stop storing data in a house. So, motion-triggered system can be useful for this case. Similarly, other issues can be considered. Apart from that, for privacy issue, one may not allow to store the video data in a room. Hence, instead of collecting video data directly, a system can convert the video data into skeleton or so. Yes, Kinect sensor or similar depth sensors can store skeleton data smartly inside a room when the illumination are not much varied. But there is a limitations of depth distance (up to a few meters only) and after that the depth maps or skeleton extractions are noisier. Therefore, data collection issue and modalities are crucial to ponder about.

### 12.3 Feature-Level Challenges

Features are the key for any recognition and analysis. Traditionally, features are explored a lot based on hand-crafted feature extraction methods, mostly on smaller datasets. However, deep learning era has immensely changed the fate of hand-crafted features in this domain. Now, are the hand-crafted successful features dead? Not exactly. In computer vision domain, optical flow is still widely explored but with modified approaches. Motion history image (MHI) and its variants are much less explored. Histogram of Oriented Gradients or similar other excellently-successful methods are much less explored now. Where to move? Can we get insights from the hand-crafted dominant features of the recent past to engulf with deep learning methods? Much less works have been professionally done in this arena so far. Feature
fusion is important. Feature dimension and interpretability are important too. In this section, these issues are discussed.

### 12.3.1 Handcrafted Versus Learning-Based Features

In recent years, it has been shown in numerous experiments that learned features significantly can outperform hand-crafted features in most activity analysis tasks. For example, in [33], the authors showed that for a large heterogeneous dataset of pedestrian images, the gender recognition performances of deep learning-based methods are far superior than traditional hand-crafted features such as Histogram of oriented gradient (HOG). While learned features also show better generalization capability, the authors found that for small homogeneous datasets, the performances do not differ much. In many application areas where large datasets are not available, such as medical imaging, we are still relying on traditional methods. Whereas, the overall trend in contactless human activity analysis is to use deep learning approaches wherever possible. Combining hand-crafted features with learning-based models is an interesting research domain. Training deep neural networks with a small training set while retaining reliable performance level is a big challenge. Techniques such as transfer learning, data augmentation, and regularization are being explored in this regard, but there are no specific framework [34]. Also, it is unclear why deep learning works so well across different domains and tasks. There has been a lot of experiment and explanation in this regard (e.g., [35–37]) but the scientific community has not fixed on one yet [38]. A better theoretical understanding of the deep neural networks is imperative to move forward.

However, for small datasets and for specific works like the analysis of rehabilitation patients, action scoring, action quality assessment, hand-crafted features can be useful and important. Skeleton-based activity analysis still require many hand-crafted features unless the dataset is large like NTU’s 60 classes or 120 classes. On the other hand, sensor-based domain has many works on smaller datasets based on various time and frequency domain-based features.

### 12.3.2 Fusing Features

Feature-level fusion is now a popular choice, especially when making decision from multiple modalities [39, 40]. However, there are many existing challenges related to fusing features. For example, [11] addressed temporal fusion and indicated that the encoding needed for fusion might lose temporal order information. Additionally, most 3D convolutional networks can only accommodate only a fixed small number of frames as input. Otherwise, they become computationally costly, which is most often inadequate for good temporal modeling [11]. In [41], the authors pointed out that apart from spatio-temporal information, a video sequence has structural information
that is not usually extracted or estimated. Simultaneous estimation of spatio-temporal and structural information and fusing them for decision making can be a good future challenge.

Finally, there is no specific guideline for feature fusion at present. Most researchers independently determine how they want to encode and fuse the features. What would be an optimum dimension of encoded feature to retain the most relevant information is debatable and not easy to determine without repetitive experiments. Hence, it is imperative to establish a proper guideline for feature fusion.

### 12.3.3 Feature Dimension and Interpretability

The dimension of features used for machine learning tasks vary widely depending on the methodology or the requirements. For hand-crafted feature-based methods, a feature vector can consist of a few hundred values and feature selection techniques are quite common to select the most representative feature. The advantage of hand-crafted features is that they are mostly intuitive and a researcher can justify the feature selection depending on the task at hand. However, for learning-based approaches, especially using deep neural networks, the higher level-feature dimension can be as large as several thousands. The most popular pretrained deep neural architectures provide between 512-D and 2096-D high level features. How to interpret these results or how to select the best representative features for these learning-based cases is a demanding research area. There are a few recent works on network pruning, compression and feature dimension reduction that addresses this issue and more [42–45].

As mentioned in [46], sensor data used for activity recognition are complex, unintelligible and noisy. Hence, it is important to be able to interpret such data in a way to identify relevant information and ignore everything else. Interpretabilty of the mechanism or output of a model is equally important. For example, the higher level features of a deep learning model may not be easily interpretable in human-level terms. Therefore, such outputs might not be convincing enough when making a critical decision such as determining whether a person has cancer or not from the MRI images. Unless the machine provides logical reasoning for its decision, it is very hard for a human doctor to blindly rely on that. Hence, the branch of interpretable machine learning is gaining popularity. In recent years, there has been a lot of interesting works on interpreting high-level features or the decisions produced by a model. One of the best examples in this arena is the concept of class activation maps (CAM) [47] which helps visualize the discriminative region of an input image that prompted a convolutional neural network to produce a certain classification result. In [48], an extension of the CAM method is introduced that is capable of handling more generic network architecture to provide ‘visual explanation’ of a CNN’s classification decision.

Complex temporal models are still uninterpretable and hence, almost all cases, these models remain mostly as black-boxes [49]. We need the knowledge and capabil-
ity to interpret any learning-based models. Otherwise, we can not decipher the power or rationality of the model’s decision, even though it can produce more correct results [46, 49–51]. In [52], Kim proposed a method that select crucial interpretable sensor signals. According to [46], a trusted method should have the ability to interpret the data, and have the capability of reasoning—which data are responsible for better recognition, and which part of data are failing to represent the action in a smarter manner. These issues are important and the future will lead us in better directions, if we can decipher the interpretability of the models.

12.4 Algorithm-Level/Architecture-Level Challenges

After sensor-based, database-related and feature-based related issues, we now cover the architecture and algorithmic issues for contactless human activity monitoring. In the above section, some of the constraints and issues related to features are overlapped with the network architecture and modeling, especially for deep learning-based approaches. This section covers: preprocessing issues, how to find a well-suited model, classical hand-crafted or deep network, transfer learning issues, real-time analysis and computing issues, learning with limited labels, learning in the presence of labeling errors/noises, unsupervised learning and related issues.

12.4.1 Preprocessing Steps

Proper preprocessing of images has been shown to improve performance [53] as well as training accuracy [54] of machine learning algorithms. Developing algorithms for data preprocessing is itself an interesting research domain. For example, efficient algorithms are needed to clean activity data captured with different sensors from real-world noises (as discussed in 12.2.1). With the increasing number of large datasets, preprocessing big data is now a challenge. Challenges in big data preprocessing may include data cleaning, missing value imputation [55], feature selection, data integration, and data normalization [56].

12.4.2 Models and System Architecture

In recent years, there has been a revolution in machine learning models with the emergence of a wide variety of deep neural network frameworks. Even though such frameworks pushed the performances of HAR systems to significantly higher levels, they also increased the variability in system design many-fold. As machine learning systems are still lagging far behind the human level, resolving the following challenges might help them get closer.
12.4.2.1 Finding the Best-Suited Model

At present, the choice of the best-suited model for an activity analysis task depends on a multitude of factors. In accordance to the ‘No Free Lunch Theorem’ [57], due to a large number of variables in play for real-world scenarios, it is nearly impossible to guarantee that a certain solution (for any machine learning task, not only contactless activity analysis) will perform equally well every situation. In [23], the authors described that for a human activity classifier, factors like feature types, size of training data, choice of the loss function, hyperparameters, and priors vary widely across models making each model work well for a specific domain/setup. In contrast, fail badly for others for the same task. Based on these facts, the authors in [23] argued that while classification accuracy is the primary concern, useful features must be extracted and provided to the classification system to simplify the classification problem itself.

12.4.2.2 Traditional Versus Neural Network-Based Solutions

Most of the earlier activity analysis approaches are based on traditional machine learning methods. However, with the advent of deep-learning-based strategies and faster computing facilities, there is a paradigm shift in feature extraction and classification. While deep neural networks have shown to be significantly accurate at modeling complicated class differences, they do require large datasets with a wide range of variations and also do not guarantee convergence to a global solution [23]. On the other hand, simple classifiers like linear support vector machine (SVM) can train a small dataset and find a global solution—but not generalize well across scenarios and wild variations. Additionally, there are the issues of computational burden, interpretability of the features, robustness against adversarial attacks, and ease of implementation that play a role in choosing between the two courses of solutions. A future challenge is to either establish a proper guideline for making the choices or consolidate the two courses into one framework theoretically to harness the advantages of both.

The concept of Quantum machine learning is gaining heavy traction recently for its exponential advantages over the classical methods [58]. The quantum algorithms can represent bigger range of probability distributions and operate in an exponentially faster speed [59]. This brand new branch of research also brings along a lot of unanswered questions. In [59] the authors mentioned a few, for example, how to determine if a problem is more suitable for quantum computer in comparison to a classical one, how to efficiently analyze large quantum data sets, how do we develop a unified quantum learning theory and so on.

12.4.2.3 Transfer Learning

Transfer learning is an intelligent solution to train deep neural networks with a small amount of labeled data. It is predicted that the trend of using transfer learning will
continue to grow in all computer vision applications, including human activity analysis [60]. In [61], a comprehensive survey on the use of transfer learning for activity recognition is presented. The authors discussed that finding a domain-independent generalizable distance measure for discriminating the source and target populations can be useful to compare and evaluate transfer learning approaches. Also, such a measure might help mitigate or at least identify the occurrences of negative transfer effects, a phenomenon when the performance drops for transfer learning due to due to task dissimilarity [61]. In [62] the authors stated three main challenges:

- There are no universally accepted unified framework for transfer learning, which creates many different formulations as well as confusion.
- Measuring knowledge gains from transfer learning, especially when the new learned classes are different from the original set, and
- Theoretical reasoning on the behavior of transfer learning approach for dissimilar datasets.

### 12.4.2.4 Combining Models

A popular method for handling multiple data modalities is to combine different types of models in the same solution framework. For example a combination of CNN and RNN-LSTM networks are frequently used for to harness spatial as well as temporal information from videos for human activity recognition [60, 63, 64]. Developing a standard protocol for training multiple models cooperatively can be a good research direction. In [11] the authors suggested fusing RGB, skeleton, and depth modalities as a direction to achieve superior HAR performance. A few such approaches have been reported very recently, and they’ve shown promising performance [65, 66].

Some human activity analysis tasks, by definition relies on multiple modalities, such as utilizing audio-visual data for tasks like emotion recognition, audio-visual separation and localization, cross-modality information synthesis, correspondence learning, and representation learning require processing synchronized speech and video sequence data [67]. Another recent work combines Wifi signal frames and videos to design the HAR system for indoor activities [68]. Clearly, the research on human activity analysis is shifting towards multi-modal frameworks.

### 12.4.3 Real-Time Motion Analysis

For any practical motion analysis system, producing output instantaneously for a continuous stream of real data is plausibly the ultimate target [11]. However, most human activity researches focus on handling a pre-segmented set of video frames, whereas continuous/online framework cannot assume action boundaries apriori [11]. Usually, a cache is used to store a small temporal window of frames that an algorithm can process and right away produce estimation. In [11], the authors mentioned two
prominent approaches for online motion recognition namely, sliding window-based [69] and Recurrent Neural Network (RNN)-based [70]. Both approaches have their pros and cons when considering practical implementation and the wide variety of design choices adds to the complexity of designing an effective system for real-time motion analysis.

12.4.4 Learning On-the-Fly

While real-time inference for any activity analysis model is a big challenge, an even bigger challenge is to enable the model to learn on-the-fly. Learning deep neural network on online setting—on-the-fly is a challenge and difficult task till-to-date for various applications [24, 71]. For the case of shallow networks, it is relatively less cumbersome to learn on online setting, as the optimizations on linear or kernel-based hypothesis are based on convex optimization function. On the other hand, for deep networks, the objective functions’ optimizations are not convex but non-convex [71]. The online learning face the challenges of convergence issues, for example, diminishing feature reuse, vanishing gradient, optimal depth of the network, scalability of the instances, etc. [71]. So far, the progress on this arena is not much, especially for larger datasets on video-based activity recognition. In the future, we need to address this issue for smarter activity analysis, on-the-fly.

12.4.5 Learning with Limited Labels

Collecting sufficient labeled training data for all sorts of activity analysis tasks might be nearly impossible in practice—for example, data on all kinds of anomalous conditions in autonomous driving or medical data for rare diseases. Learning with very few labeled samples, known as ‘few-shot learning,’ is a popular and challenging research topic at present [72–75]. Especially, there is no clear guideline yet on how to effectively make deep learning frameworks to learn from few labeled samples. Along this direction, cross-domain few-shot learning is an even more challenging scenario where the source and target domains are dissimilar. Yet, the few-shot transfer learning is to be done for the target domain [72]. The evaluation benchmarks for cross-domain few-shot learning are still not well-defined and [72] is one of the few studies that tried to bridge the gap.

12.4.6 Learning in the Presence of Label Noise

Labeling noise is bound to be present in any manually annotated very large dataset. It is known as one of the critical factors affecting the performances of CHAA. For example, In [76] the authors showed that for a face recognition task, a model trained
with a subset of the MegaFace dataset containing manually cleaned 32% images can reach the same performance level of training with the full dataset, demonstrating that a few order labeled data is required for face recognition if label noise is present. Automatic detection and correction of label noise, modeling label noise and learning from multiple noisy annotations are an exciting challenge which is receiving more attention in recent times as the dataset are growing in sizes [77–81]. Unsupervised learning is another way to bypass the labeling noise issue. However, the top-performing unsupervised systems are yet to catch-up with the best-supervised methods.

### 12.4.7 Reduction of Multiply-Adds and Parameters of Networks

Reduction of network axes in terms of time, depth and width are important for faster processing. When we explore higher resolution (spatio-temporal), due to the higher number of parameters, more multiplication and addition steps, the network becomes costly in terms of computation. There should have a trade-off between globally expanding the network width and expanding the inner bottleneck width [82]. This trade-off has been considered for efficient CNN for embedded hardware designs and mobile architectures [83, 84]. Balance between the network resolution and channel capacity is another important point to explore.

### 12.4.8 Unsupervised Learning

As mentioned earlier, with the increasing size of the human activity dataset and the associated cost of annotating manually, it is imperative to develop the machine learning frameworks for efficient and more precise unsupervised learning. To be able to utilize the massive bulk of data being gathered every day without human intervention is one of the biggest future challenges for the human activity research community. Especially considering the wide variation in human activities and scenarios [11, 60]. Recently, in [85], the authors presented an unsupervised learning framework for mobile robots in a real-world scenario. The framework utilizes the probabilistic Latent Dirichlet Allocation (LDA) technique to bypass the need for manual temporal segmentation of videos and models each observation as a “probabilistic mixture over an underlying number of latent topics, where some topics can be considered “interesting” human activities” [85]. However, when compared with supervised methods, unsupervised learning-based techniques are lagging way behind as of now.
12.5 Implementation-Level Challenges

When it comes to implementing a human activity analysis system for any practical purposes, some important issues are reaching the best achievable performance level, identifying the failure cases, and finding the optimum operating point at which the system would work reliably [7].

12.5.1 Robust Performance

The system should perform robustly despite changes in lighting, pose variations or partially occluded human bodies, and background clutter [7]. In [23], the authors pointed out that most research works only provide overall performance results and do not provide insightful results to understand the robustness of the methods against viewpoints, occlusion, pose and illumination variation, and background clutter [86].

12.5.2 Real-Time and On-the-Edge Solution

High computational complexity is a known challenge for vision-based activity analysis tasks, e.g., skeleton-based human activity recognition [24]. In many practical applications such as security systems, anomaly and fall detection systems, authentication real-time or near real-time understanding of activities are of paramount importance. Huge research efforts are underway in academia and industry to achieve near real-time solutions for very complex systems so that such systems can be deployed in smartphones, extended reality glasses, and smart appliances. Network compression by parameter pruning, low-bit quantization, knowledge distillation, and designing special structural convolutional filters are interesting new research areas in this domain [87]. Also, many commercial solutions are relying on distributing the workload between edge devices and cloud to ensure smooth performance. This requires designing feasible solutions for the edge devices as well as boosting edge computing through innovation. One such innovative approach is the concept of federated learning that aims at privacy preserving machine learning that significantly reduces high communication cost [88–90].

---

1 https://venturebeat.com/2020/07/28/nvidia-bmw-red-hat-and-more-on-the-promise-of-ai-edge-computing-and-computer-vision/.
12.5.3 Low Power, High-Performance Solutions

The most desirable activity analysis solutions are to be low power, low latency with robust performance. Especially, solutions designed for smartphones or extended reality headsets need to be power-efficient yet real-time and high performance. Apart from model compression, a few other factors are considered for designing efficient solutions, such as input resolution for visual sensors, neural network architecture, sampling rate, temporal and spatial window sizes, and trade-offs between wide versus deeper models. At present, there is no clear guideline on designing a low power high-performance solution in a systematic way, which usually leads to intuition-based design changes followed by multiple trial and error to find a feasible solution.

12.5.4 Benchmark and Performance Evaluation

Different research work on the same problem domain reporting results using different performance metrics and on different benchmark datasets makes comparative evaluation nearly impossible. Also, variation in experimental protocol sometimes adds more complexity. There is a clear lack of a standard for experimental protocol, consolidated benchmark, and evaluation metrics that is acceptable to all the activity analysis researchers worldwide. Such standards must be established soon to ensure measurable and comparable advances in the contactless activity analysis domain.

12.5.5 Domain Invariant CHAA

Due to the large variance in distribution between different datasets targeted for different action recognition tasks performed in various scenarios and captured using different types of sensors, it is a challenge to find a domain invariant solution that would work across datasets [11]. Transfer learning techniques are currently being used to quickly learn about a new domain, but it requires sufficient labeled data on the target domain. Recently, an exciting work [91] used transfer-learning to an unlabeled target domain by unsupervised domain adaptation (UDA). The proposed approach uses adversarial alignment and multi-modal self-supervision to update fine-grained action recognition models for unlabeled target domain video sequences. Future research in this area can focus on multi-modal fusion and multi-task self-supervision [91].
12.5.6 Multi-view Human Activity Analysis

Research on multi-view human activity recognition and analysis has been gaining traction in recent years and it is one of the most difficult yet essential challenge to be solved [92, 93]. In [93], the authors divided the multi-view human action recognition methods into four categories, namely, geometric constraints, view-invariant features, human body joint tracking, and knowledge transferring. The research direction on this challenge include accurate 3D human pose estimation from multi-view camera setups [94], learning view-invariant spatio-temporal features [95], feature fusion for view-invariance [96], sparse view-invariant representation [97], 3D human skeleton tracking [98, 99] and knowledge transfer across views to achieve view-invariance [100].

12.6 Application-Level Challenges

New application areas requiring human activity analysis are emerging nowadays, fueled by the rapid advancement of theoretical understanding of the physical world, computational capability, and the growing popularity of such applications in daily living. The impact of having robust human activity analysis systems is far-reaching, influencing government policies, social behavior [101], security, trade, healthcare [102], and even nature [103]. A few emerging application areas and associated challenges are discussed below:

12.6.1 Surveillance

Surveillance systems can be of many forms ranging from a complex large-scale crowd monitoring to a simple baby-monitoring system. There is a growing consumer market for intelligent surveillance system backed by AI and machine learning. In this subsection, we touch base on a few of the trending use-cases of such systems.

12.6.1.1 Mass-Scale Crowd Monitoring

A huge emerging application area of contactless activity analysis is large scale crowd surveillance. Challenging use cases to include detection of anomalous situations, understanding social behavior, finding correlations between actions performed by different entities, and recognizing groups of subjects that are socially engaged [24]. Newer use cases are also being introduced regularly. For example, recently, mass-scale crowd monitoring is being considered as a measure against the COVID-19

2https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847.
pandemic to ensure physical distancing between individuals to prevent the spread of Coronavirus [104].

### 12.6.1.2 Surveillance Using Unmanned Aerial Vehicles (UAVs)

An integral part of designing fully autonomous search and rescue UAVs to support during a disaster, war, and many other crises are to detect and even identify human under distress [105]. Detecting humans, analyzing and recognizing human activity from aerial imagery robustly in real-time is a big future challenge.

### 12.6.1.3 Advanced Home Surveillance

Human activity analysis, tracking, and recognition are integral parts of intelligent home surveillance systems. Reliable surveillance systems require smart scene segmentation, object detection and tracking, feature analysis, context modeling and occlusion handling capabilities [106]. Video-based contactless residential home surveillance systems have major privacy concerns that needs requires some innovative solution [107].

Multi-view and multi-sensor systems will dominate the field in the future and fusing different sensor data to achieve the most reliable performance is going to be a big challenge. Apart from sensor data, a few other factors also need a closer look. For example, a clear guideline is needed for the interaction between all the intelligent components of an advanced home surveillance system so that one component such as a drone surveilling the nearby street can confirm with a smart lawn camera about a possible break-in. At present, lack of realistic benchmark datasets in this domain has made is difficult to realize and evaluate viable frameworks. Hence, collection and annotation of challenging real-world scenarios for advanced home surveillance is a necessity.

### 12.6.1.4 Automated Health Surveillance

In the healthcare industry, automated surveillance is going to bring a huge shift. Fall detection of elderly individuals [108, 109], indoor surveillance for health hazards and anomalous situations [110], remote monitoring of vitals [111, 112], predictive modeling to prevent accidents, surveillance of social behavior to predict mental health issues [113, 114], and mass-scale surveillance to prevent infectious disease propagation [115, 116] are going to be part of human lives in not-so-far future. Contactless health surveillance is being heavily used as a measure to prevent the spread of coronavirus. For example, the airport security protocol now includes checking vitals of the passengers in a contactless manner ³Remote health monitoring is also used in neonat-

³https://airport-world.com/australias-avalon-airport-installs-pioneering-touchless-technology/.
tal intensive care units (NICU) [117] and to monitor respiration and heart rates of severely burned patients [112]. In the future, automated health surveillance systems may become an integral part of any surveillance system - be that a small scale home surveillance or large scale population surveillance framework. The system should be able to do estimate the vitals of the subjects in the frame and detect anomalies in physiological or behavioral patterns. For monitoring a single person, such kind of systems are already being deployed for smart cars, where the central AI of the car continuously monitors the driver to detect drowsiness or fatigue and generates a warning signal accordingly [118–121]. Major challenges of contactless automated health surveillance include robust detection and separation of the source/subject, retaining high precision and recall, invariance to environmental noise and sample biases, and effective feature-level or score-level fusion of information from multiple sensors. For most real-world applications, a contactless automated health surveillance system needs to be light-weight while capable of utilizing multi-sensor input streams efficiently to give very accurate results. It is indeed a huge challenge.

### 12.6.2 Extended Reality Applications

In most extended reality (XR) applications, continuous detection, and tracking of the user’s hands using only vision sensors placed on the head-mounted devices are gaining traction. Such technology is vital for applications like controller-free gaming and fully virtual workstations. Popular virtual and augmented reality solutions such as Oculus Quest and Microsoft Hololens are already working towards seamless hand detection, tracking, and hand gesture recognition to ensure the best user experience. Real-time, robust hand detection, tracking and hand gesture recognition from multi-view and multi-modal camera systems is a big challenge to overcome to make XR technologies more appealing to the mass population.

### 12.6.3 Action Prediction

For many practical use cases, action prediction is more important than recognition. For example, pedestrian and driver behavior predictions are vital for safely operating autonomous or smart vehicles [11, 122]. Also, predictive modeling of an anomaly, a possible criminal activity, the behavior of a crowd, or a mental health issue can be very useful [23]. The question is, how reliably can we predict future actions based on the past and current scenarios? The task is very challenging.

---

4https://ai.facebook.com/blog/hand-tracking-deep-neural-networks.
5https://docs.microsoft.com/en-us/hololens/hololens2-basic-usage.
12.6.4 Action Localization

Isolating the locations with higher actionness scores is a difficult task, but it can ease the process of recognition. At the same time, it is a challenge to remove or give less importance to non-action shots. Spatio-temporal localization of action from a sequence of frames is a very difficult yet useful task [23]. Considering multiple modalities such as audio and video might provide better localization cues. There are recent works on learning temporal co-attention models for unsupervised action localization [123]. Considering a context-aware loss is also found to be useful for such tasks [124]. The domain has a lot of future research potential.

12.6.5 Game-Play Analysis

Game-play analysis helps game designers understand the users’ behavior in a game and improve the design by eliminating the chances of unexpected behavior. Analysis of digital game that is being played either locally or in a server is an interesting future research direction. Especially, a robust solution will help automate the process and may reduce the time and cost of game-analysis significantly. In a recent work [125] the authors presented a game-play analysis approach for first-person shooter games by only using the visual information. The robustness of such systems can be improved by incorporating information from other sensors such as audio and multiple views.

12.6.6 Imitation Learning or Action Mimicking

Imitation learning to train robots perform actions similar to humans is a dominant research field at present [126, 127]. Imitation learning is most useful in the cases when demonstrating a task repetitively is easier for an expert than defining a reward function, and a machine can learn a policy from that demonstration to mimic the task performed [126]. Such learning methodology can be used in many AI applications, including remote robotic surgeons, human-computer interaction, self-driving vehicles, and computer gaming [126]. The two main techniques for imitation learning are behavioral cloning and inverse reinforcement learning [127]. Embodiment mismatch and viewpoint difference are two of the main challenges faced by vision-based imitation learning techniques. Even though the topic has been actively researched for the past two decades [128, 129], the potentials of many recent machine learning techniques such as adversarial learning and domain adaptation are yet to be explored in this domain [127].
12.6.7 Assistive Technology

As discussed in the 2017 Consensur Report from the National Academies of Sciences, Engineering, and Medicine [130], by availing appropriate assistive products and technologies and through proper training while limiting societal and environmental barriers it is possible to partially or completely mitigating the impacts of impairments and enhancing work participation. Contactless activity monitoring is expected to play a crucial role in taking the assistive technologies for the physically and mentally challenged to the next level. Smart devices such as smart augmented reality (AR) glasses and head mounted virtual reality (VR) devices will help people from all aspects of life to improve their quality of living. For example, AR glasses and VR headsets will help builders to find structural defects and engineers to analyze a machine quickly by overlaying useful information on the objects in the screen. AR glasses will assist soldiers to know their surroundings, help athletes and pilots alike to train in a more realistic virtual environment, and assist mentally and physically impaired to learn new skills. By analyzing the scene and the persons hand movement, eye gaze and head rotation, the AR and VR devices are going to enable the users to interact and navigate in the digital environment in a seamless fashion. Apart from the AR and VR technologies, the field of assistive technologies is enjoying a rapid growth in terms of concepts, available products and usability through innovation. Some of the most popular use cases of current and future technologies that would rely on contactless human activity analysis are

- Interpreting signs and language [131–133]
- Navigation assistance such as smart guiding glasses [134]
- Interpreting and conveying information on scenery, aesthetics, or beauty for visually impaired [135–137]
- Gesture or voice controlled smart-home systems such as Amazon Echo, Google Home or Apple HomeKit [138]
- Contactless health-monitoring [110, 112, 116]
- Virtual, Augmented and Extended reality assistance for people with physical and mental disabilities [139–144]
- Virtual reality worlds and Virtual events where an avatar mimics the users expression and action [145]

12.6.8 High-Level Human-Computer Interaction

In [146], a group of 32 researchers from the HCI domain presented seven grand challenges related to HCI. The challenges are:

1. Human-Technology Symbiosis
2. Human-Environment Interactions
3. Ethics, Privacy, and Security
4. Well-being, Health, and Education
5. Accessibility and Universal Access
6. Learning and Creativity
7. Social organization and democracy.

Most, if not all, of these challenges are related to human activity analysis one way or another. For example, one aspect of human-technology symbiosis is the adaptation and personalization of human needs, including detecting and predicting human activities/intentions in a non-invasive manner [146]. In human-environment interaction, a futuristic concept is escalated interactions through smell, taste, and haptic sensations in addition to vision, hearing, and touch, which are already being explored heavily. Developing a more realistic feeling of immersion and presence in the extended reality (XR) environments is another topic under this challenge. Ethical use of HCI research outcomes while ensuring privacy and security is a major concern. For example, human activity analysis can extract private information from a scene when tracking a subject’s every move—it is imperative to ensure that proper data abstraction and encryption are in place to prevent information leakage and security breach. The use of HCI in well-being, health, and education involves interpreting human activities. One interesting example presented in [146] is the ‘Serious Games’ played to drive health-related outcomes. The games can motivate and engage a user to participate in active physical exercise by providing intelligent feedback on the current state and future goals—all of which can involve human activity analysis. Prevention, compensation, and support for older people through ambient assisted living is another challenging domain under healthcare using HCI. The development of such assistive technologies is also part of ensuring universal access to technology. Next, an immersive learning approach using XR can bring a paradigm shift in traditional learning approaches and insight creativity in new directions. Seamless tracking of hand and body movements to interpret the interaction between the physical and digital world will play a vital role in implementing such approaches. Finally, the challenges related to social organization and democracy involve developing technologies for crisis response (e.g., during natural disasters), deficient healthcare, and unbiased social justice [146]. Contactless human activity analysis will definitely play a major role in materializing these goals.

12.7 Conclusion

Research on contactless human monitoring has gained unprecedented traction in recent years for a multitude of co-occurring phenomena such as the rapid growth of processing technology, demonstrated effectiveness of deep learning-based methods, advancement in sensor technologies, and an unforeseen increase in data storage capabilities. This chapter touched base with the most challenging present issues pertaining to contactless human activity analysis and tried to interpolate the current research works in the future direction. To give the readers a holistic idea about the true nature of the challenges, we segmented human activity monitoring tasks into five...
elemental components, namely, sensors, features, algorithm, implementation, and applications. Our discussion on each component ranged from traditional challenges from the past to the most advanced issues addressed in the literature at present. Furthermore, we tried to shed light on the future based on an extensive literature search and putting our thoughts and ideas in the mix.

Achieving low power, low bandwidth while retaining high resolution is already a top-priority sensor-level challenge. Multi-sensor system design and privacy issues are expected to climb up the priority list rapidly over time. For the feature-level, learning-based features are already replacing hand-crafted features. Feature fusion and interpretability will probably continue to dominate research in this arena for some time. Among the algorithm-level challenges, finding a systematic way to choose the most appropriate framework or developing a scalable unified framework for the activity monitoring task is paramount. Also, an advanced understanding of transfer learning and achieving superior learning-on-the-fly capabilities with limited labels data or in an unsupervised fashion can be expected to dominate future research. The most notable implementation-level challenges would be to develop a robust, real-time, on-the-edge solution capable of operation in different domains and handle multiple sensory inputs. Finally, the application-level challenges are found to be varying widely depending on use cases. Some dominating future use cases would be high accuracy mass-scale crowd monitoring system, automated non-intrusive health surveillance, a merge of extended reality in day-to-day life, imitation learning for HCI, and developing advanced assistive technology.

In recent years, we are experiencing a revolution in the number, topic, and depth of machine learning research on contactless human activity monitoring and analysis. Over time this trend will only grow, and we believe the future of research in this domain is brighter and more exciting than ever before. Our daily lives are being a little more automated every day, and we, the humans, are incessantly adapting to the new technological norms. Human activity analysis plays probably the most pivotal role in this transformation. Hence, a substantial idea of current and future challenges is essential for every researcher in this field. We believe, persistent and innovative research on contactless human activity analysis and thoughtful usage of the outcomes will help shape the future for the betterment of humanity.

References

1. Poppe, R.: A survey on vision-based human action recognition. Image Vision Comput. 28(6), 976–990 (2010)
2. Ahad, M.A.R.: Vision and sensor based human activity recognition: Challenges ahead, chapter 2. In: Advancements in Instrumentation and Control in Applied System Applications, IGI Global, pp. 17–35 (2020)
3. Ahad, M. A. R., Lago, P., Inoue, S.: Human activity recognition challenge, Publisher: Springer Nature Switzerland AG (2020)
4. Ahad, M.A.R.: Computer vision and action recognition: A guide for image processing and computer vision community for action understanding, ISBN: 978-94-91216-20-6, available
5. Rodríguez, N.D., Cuéllar, M.P., Lilis, J., Calvo-Flores, M.D.: A survey on ontologies for human behavior recognition. ACM Comput. Surv. 46(4) (2014)

6. Akdemir, U., Turaga, P., Chellappa, R.: An ontology based approach for activity recognition from video. In: Proceedings of the 16th ACM International Conference on Multimedia, MM ’08, pp. 709–712, New York, NY, USA, Association for Computing Machinery (2008)

7. Vrigkas, M., Nikou, C., Kakadiaris, I.A.: A review of human activity recognition methods. Front. Robot. AI 2, 28 (2015)

8. Perry, S.: Image and Video Noise: An Industry Perspective, pp. 207–234. Springer International Publishing, Cham (2018)

9. Yang, X., Zhang, Z.: Reconstruction of underwater images with distortion using robust image registration. In: Asundi, A., Fujigaki, M., Xie, H., Zhang, Q., Zhang, S., Zhu, J., Kemao, Q. (eds.) Seventh International Conference on Optical and Photonic Engineering (icOPEN 2019), vol. 11205, pp. 116–121. International Society for Optics and Photonics, SPIE (2019)

10. Chellappa, R.,rgios Theodoridis, (eds.) Chapter 1—Multiview Video: Acquisition, Processing, Compression, and Virtual View Rendering, pp. 3–74. Academic Press (2018)

11. Wang, P., Li, W., Ogunbona, P., Wan, J., Escalera, S.: RGB-D-based human motion recognition with deep learning: a survey. Comput. Vis. Image Underst. 171, 118–139 (2018)

12. Chen, F., Masi, C.: Effect of noise on automatic speech recognition system error rate. Proc. Hum. Factors Ergon. Soc. Ann. Meet. 44(37), 606–609 (2000)

13. Guang Yi Chen: An experimental study for the effects of noise on face recognition algorithms under varying illumination. Multi. Tools Appl. 78(18), 26615–26631 (2019). Sep

14. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725–1732 (2014)

15. Ranjan, R., Sankaranarayanan, S., Bansal, A., Bodla, N., Chen, J., Patel, V.M., Castillo, C.D., Chellappa, R.: Deep learning for understanding faces: Machines may be just as good, or better, than humans. IEEE Signal Process. Mag. 35(1), 66–83 (2018)

16. Ahad, M.A.R., Ngo, T.T., Antar, A.D., Ahmed, M., Hossain, T., Muramatsu, D., Makihara, Y., Inoue, S., Yagi, Y.: Wearable sensor-based gait analysis for age and gender estimation. Sensors 20(8), 2424 (2020)

17. Mahbub, U., Sarkar, S., Patel, V.M., Chellappa, R.: Active user authentication for smartphones: a challenge data set and benchmark results. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–8 (2016)

18. Mahbub, U., Sarkar, S., Chellappa, R.: Partial face detection in the mobile domain. Image Vision Comput. 82, 1–17 (2019), ISSN 0262-8856. https://doi.org/10.1016/j.imavis.2018.12.003

19. Mahbub, U., Komulainen, J., Ferreira, D., Chellappa, R.: Continuous authentication of smartphones based on application usage. In: IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 1, no. 3, pp. 165–180, July 2019. https://doi.org/10.1109/TBIOIM.2019.2918307

20. Mahbub, U., Chellappa, R.: PATH: Person authentication using trace histories. In: 2016 IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, 2016, pp. 1-8, https://doi.org/10.1109/UEMCON.2016.7777911

21. Papamichail, M.D., Chatzidimitriou, K.C., Karanikiotis, T., Oikonomou, N.-C.I., Symeonidis, A.L., Saripalle, S.K.: Behavioral biometrics dataset towards continuous implicit authentication, March 2019

22. Sitova, Z., Sedenka, J., Yang, Q., Peng, G., Zhou, G., Gasti, P., Balagani, K.S.: Hmog: New behavioral biometric features for continuous authentication of smartphone users. IEEE Trans. Inf. Forensics Secur. 11, 877–892 (2016)

23. Kang, S.-M., Wildes, R.P.: Review of action recognition and detection methods. ArXiv, arXiv:1610.06906 (2016)

24. Presti, L.L., La Cascia, M.: 3D skeleton-based human action classification: a survey. Pattern Recognit. 53, 130–147 (2016)
25. Zeng, M., Yu, T., Wang, X., Nguyen, L.T., Mengshoel, O.J., Lane, I.: Semi-supervised convolutional neural networks for human activity recognition. In: 2017 IEEE International Conference on Big Data (Big Data), pp. 522–529 (2017)

26. Mabrouk, M.F., Ghanem, N.M., Ismail, M.A.: Semi-supervised learning for human activity recognition using depth cameras. In: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 681–686 (2015)

27. Flouty, E., Zisimopoulos, O., Stoyanov, D.: Faceoff: anonymizing videos in the operating rooms. In: Stoyanov, D. (eds.), OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, pp. 30–38. Springer International Publishing, Cham (2018)

28. Dai, J., Wu, J., Saghaei, B., Konrad, J., Ishwar, P.: Towards privacy-preserving activity recognition using extremely low temporal and spatial resolution cameras. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 68–76 (2015)

29. Dwork, C., Roth, A.: 2014. The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. 9, 3–4 (August 2014), 211–407. https://doi.org/10.1561/0400000042

30. Oculus. Hand tracking

31. Bo, N.B., Deboeverie, F., Eldib, M., Guan, J., Xie, X., Niño, J., Van Haerenborgh, D., Slembrouck, M., Van de Velde, S., Steendam, H., Veeelaert, P.: Human mobility monitoring in very low resolution visual sensor network. Sensors (Basel, Switzerland) 14(11), 20800–20824 (2014)

32. Xiwei Huang, Yu., Liu, J.X., Hang, X., Han, Z., Rong, H., Yang, H., Yan, M., Hao, Yu.: Machine learning based single-frame super-resolution processing for lensless blood cell counting. Sensors (Basel, Switzerland) 16(11), 1836 (2016). Nov

33. Antipov, G., Berrani, S.A., Ruchaud, N., Dugelay, J.L.: Learned vs. hand-crafted features for pedestrian gender recognition. In: Proceedings of the 23rd ACM International Conference on Multimedia, MM ’15, pp. 1263–1266, New York, NY, USA, Association for Computing Machinery (2015)

34. Brigato, L., Iocchi, L.: A close look at deep learning with small data (2020)

35. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning requires rethinking generalization. In: The 5th International Conference on Learning Representations (2017) Cite arxiv:1611.03530Comment: Published in ICLR 2017

36. Lin, H.W., Tegmark, M., Rolnick, D.: Why does deep and cheap learning work so well? J. Stat. Phys. 168(6), 1223–1247 (2017). Sep

37. Ackley, D.H., Hinton, G.E., Sejnowski, T.J.: A learning algorithm for Boltzmann machines. Cogn. Sci. 9(1), 147–169 (1985)

38. Sejnowski, T.J.: The unreasonable effectiveness of deep learning in artificial intelligence. Proc. Natl. Acad. Sci. (2020)

39. Castro, F.M., Marín-Jiménez, M.J., Guil, N., de la Blanca, P.N.: Multimodal feature fusion for CNN-based gait recognition: an empirical comparison. Neural Comput. Appl. (2020)

40. Wang, M., Tighe, J., Modolo, D.: Combining detection and tracking for human pose estimation in videos. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2020

41. Jain, A., Zamir, A.R., Savarese, S., Saxena, A.: Structural-RNN: deep learning on spatio-temporal graphs. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5308–5317 (2016)

42. Zhao, R., Hu, Y., Dotzel, J., De Sa, C., Zhang, Z.: Improving neural network quantization without retraining using outlier channel splitting. In: International Conference on Machine Learning (ICML), pp. 7543–7552, June 2019

43. Qin, H., Gong, R., Liu, X., Shen, M., Wei, Z., Yu, F., Song, J.: Forward and backward information retention for accurate binary neural networks. In: IEEE CVPR (2020)

44. Chen, H., Wang, Y., Xu, C., Yang, Z., Liu, C., Shi, B., Xu, C., Xu, C., Tian, Q.: Data-free learning of student networks. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 3513–3521 (2019)
45. Gomez, A.N., Zhang, I., Kamalakara, S.R., Madaan, D., Swersky, K., Gal, Y., Hinton, G.E.: Learning sparse networks using targeted dropout. *ArXiv*, arXiv:1905.13678 (2019)

46. Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., Liu, Y.: Deep learning for sensor-based human activity recognition: overview, challenges and opportunities. *ArXiv*, arXiv:2001.07416 (2020)

47. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: CVPR (2016)

48. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626 (2017)

49. Kim, T.S., Reiter, A.: Interpretable 3D human action analysis with temporal convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1623–1631 (2017)

50. Ribeiro, M.T., Singh, S., Guestrin, C.: Why should I trust you?: explaining the predictions of any classifier (2016)

51. Lipton, Z.C.: The mythos of model interpretability (2017)

52. Kim, E.: Interpretable and accurate neural networks for human activity recognition. *IEEE Trans. Industr. Inf.* 16(11), 7190–7198 (2020)

53. Pal, K.K., Sudeep, K.S.: Preprocessing for image classification by convolutional neural networks. In: 2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), pp. 1778–1781 (2016)

54. LeCun, Y., Bottou, L., Orr, G.B., MÜller, K.R.: Efficient backprop. In: Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop, pp. 9–50. Springer-Verlag, Berlin, Heidelberg (1998)

55. Hossain, T., Ahad, M.A.R., Inoue, S.: A method for sensor-based activity recognition in missing data scenario. *Sensors* 20(14), 3811 (2020)

56. García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M., Herrera, F.: Big data preprocessing: methods and prospects. *Big Data Anal.* 1(1), 9 (2016)

57. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* 1(1), 67–82 (1997)

58. Dunjko, V., Wittek, P.: A non-review of Quantum Machine Learning: trends and explorations. *Quant. Views* 4, 32 (2020). March

59. Sarma, S., Deng, D.L., Duan, L.M.: Machine learning meets quantum physics. *Phys. Today* 72, 48–54 (2019)

60. Pham, H.H., Khoudour, L., Crouzil, A., Zegers, P., Velastin Carroza, S.A.: Video-based human action recognition using deep learning: a review (2015)

61. Cook, D., Feuz, K.D., Krishnan, N.C.: Transfer learning for activity recognition: a survey. *Knowl. Inf. Syst.* 36(3), 537–556 (2013). Sep

62. Sousa, R., Silva, L.M., Alexandre, L.A., Santos, J., De Sá, J.M.: Transfer learning: current status, trends and challenges. In: 20th Portuguese Conference on Pattern Recognition (2014)

63. Ahmad, W., Kazmi, B.M., Ali, H.: Human activity recognition using multi-head CNN followed by LSTM. In: 2019 15th International Conference on Emerging Technologies (ICET), pp. 1–6 (2019)

64. Mutegeki, R., Han, D.S.: A CNN-LSTM approach to human activity recognition. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC), pp. 362–366 (2020)

65. Ehatisham-Ul-Haq, M., Javed, A., Azam, M.A., Malik, H.M.A., Irtaza, A., Lee, I.H., Mahmood, M.T.: Robust human activity recognition using multimodal feature-level fusion. *IEEE Access* 7, 60736–60751 (2019)

66. Khaire, P., Imran, J., Kumar, P.: Human activity recognition by fusion of RGB, depth, and skeletal data. In: Chaudhuri, B.B., Kankanhalli, M.S., Balasubramanian Raman, M.S. (eds.), Proceedings of 2nd International Conference on Computer Vision & Image Processing, pp. 409–421. Springer, Singapore (2018)

67. Khaire, P., Imran, J., Kumar, P.: Deep audio-visual learning: a survey. *ArXiv*, arXiv:2001.04758 (2020)
68. Zou, H., Yang, J., Das, H.P., Liu, H., Zhou, Y., Spanos, C.J.: Wifi and vision multimodal learning for accurate and robust device-free human activity recognition. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 426–433 (2019)

69. Liu, C., Hu, Y., Li, Y., Song, S., Liu, J.: Pku-mmd: a large scale benchmark for continuous multi-modal human action understanding. In: ACM Multimedia Workshop (2017)

70. Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., Kautz, J.: Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4207–4215 (2016)

71. Sahoo, D., Pham, Q., Lu, J., Hoi, S.C.: Online deep learning: learning deep neural networks on the fly (2017)

72. Guo, Y., Codella, N.C., Karlinsky, L., Codella, J.V., Smith, J.R., Saenko, K., Rosing, T., Feris, R.: A broader study of cross-domain few-shot learning (2019)

73. Chen, W.Y., Liu, Y.C., Kira, Z., Wang, Y.C.F., Huang, J.B.: A closer look at few-shot classification. In: International Conference on Learning Representations (2019)

74. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: relation network for few-shot learning. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1199–1208 (2018)

75. Fei-Fei, L., Fergus, R. and Perona, P.: One-shot learning of object categories. IEEE Trans. Pattern Anal. Mach. Intell. 28(4), 594–611 (2006)

76. Wang, F., Chen, L., Li, C., Huang, S., Chen, Y., Qian, C., Change Loy, C.: The devil of face recognition is in the noise. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.), Computer Vision–ECCV 2018, pp. 780–795. Springer International Publishing, Cham (2018)

77. Arazo, E., Ortego, D., Albert, P., O’Connor, N.E., McGuinness, K.: Unsupervised label noise modeling and loss correction. In: International Conference on Machine Learning (ICML), June 2019

78. Tanno, R., Saeedi, A., Sankaranarayanan, S., Alexander, D.C., Silberman, N.: Learning from noisy labels by regularized estimation of annotator confusion. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11236–11245 (2019)

79. Zhang, Y., Deng, W., Wang, M., Hu, J., Li, X., Zhao, D., Wen, D.: Global-local GCN: Large-scale label noise cleansing for face recognition. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020

80. Shen, Y., Ji, R., Chen, Z., Hong, X., Zheng, F., Liu, J., Xu, M., Tian, Q.: Noise-aware fully webly supervised object detection. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)

81. Zhong, J.X., Li, N., Kong, W., Liu, S., Li, T.H., Li, G.: Graph convolutional label noise cleaner: train a plug-and-play action classifier for anomaly detection. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019)

82. Feichtenhofer, C.: X3d: expanding architectures for efficient video recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020)

83. Zhang, X., Zhou, X., Lin, M., Sun, J.: An extremely efficient convolutional neural network for mobile devices, Shufflenet (2017)

84. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: inverted residuals and linear bottlenecks (2019)

85. Duckworth, P., Hogg, D.C., Cohn, A.G.: Unsupervised human activity analysis for intelligent mobile robots. Artif. Intell. 270, 67–92 (2019)

86. Uddin, M.Z., Muramatsu, D., Noriko, T., Ahad, M.A.R., Yagi, Y.: Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion. IPSJ Trans. Comput. Vis. Appl. 11(9), 1–18 (2019)

87. Deng, B.L., Li, G., Han, S., Shi, L., Xie, Y.: Model compression and hardware acceleration for neural networks: a comprehensive survey. Proc. IEEE 108(4), 485–532 (2020)

88. Sozinov, K., Vlassov, V., Girdzijauskas, S.: Human activity recognition using federated learning. In: 2018 IEEE International Conference on Parallel Distributed Processing with Applications, Ubiquitous Computing Communications, Big Data Cloud
89. Kaissis, G.A., Makowski, M.R., Rückert, D., Braren, R.F.: Secure, privacy-preserving and federated machine learning in medical imaging. Nat. Mach. Intell. 2(6), 305–311 (2020). Jun
90. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Singh, A., Zhu, J. (eds.), Proceedings of Machine Learning Research, vol. 54, pp. 1273–1282, Fort Lauderdale, FL, USA, 20–22 April 2017. PMLR
91. Munro, J., Damen, D.: Multi-modal domain adaptation for fine-grained action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020
92. Wu, D., Sharma, N., Blumenstein, M.: Recent advances in video-based human action recognition using deep learning: a review. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 2865–2872 (2017)
93. Trong, N.P., Minh, A.T., Nguyen, H., Kazunori, K., Le Hoai, B.: A survey about view-invariant human action recognition. In: 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), pp. 699–704 (2017)
94. Kocabas, M., Karagoz, S., Akbas, E.: Self-supervised learning of 3d human pose using multi-view geometry. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019
95. Ji, X.-F., Qian-Qian, W., Zhao-Jie, J., Wang, Y.-Y.: Study of human action recognition based on improved spatio-temporal features. Int. J. Autom. Comput. 11(5), 500–509 (2014). Oct
96. Dhiman, C., Vishwakarma, D.K.: View-invariant deep architecture for human action recognition using two-stream motion and shape temporal dynamics. IEEE Trans. Image Process. 29, 3835–3844 (2020)
97. Zheng, J., Jiang, Z.: Learning view-invariant sparse representations for cross-view action recognition. In: 2013 IEEE International Conference on Computer Vision, pp. 3176–3183 (2013)
98. Xia, L., Chen, C.C., Aggarwal, J.K.: View invariant human action recognition using histograms of 3d joints. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 20–27 (2012)
99. Liu, M., Liu, H., Chen, C.: Enhanced skeleton visualization for view invariant human action recognition. Pattern Recogn. 68, 346–362 (2017)
100. Rahmani, H., Mian, A.: Learning a non-linear knowledge transfer model for cross-view action recognition. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2458–2466 (2015)
101. Castelluccia, C., Le Métayer Inria, D.: Impact analysis of facial recognition. Working Paper or Preprint, February 2020
102. Ahad, M.A.R., Antar, A.D., Shahid, O.: Vision-based action understanding for assistive healthcare: A short review. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, USA, pp. 1–11 (2019)
103. Fang, F., Nguyen, T.H., Pickles, R., Lam, W.Y., Clements, G.R., An, B., Singh, A., Tambe, M., Lemieux, A.: Deploying paws: field optimization of the protection assistant for wildlife security. In: AAAI (2016)
104. Punn, N.S., Sonbhadra, S.K., Agarwal, S.: Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques. ArXiv, arXiv:2005.01385 (2020)
105. Lygouras, E., Santavas, N., Taitzoglou, A., Tarchanidis, K., Mitropoulos, A., Gasteratos, A.: Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations. Sensors 19(16), 3542 (2019). Aug
106. Rho, S., Min, G., Chen, W.: Advanced issues in artificial intelligence and pattern recognition for intelligent surveillance system in smart home environment. Eng. Appl. Artif. Intell. 25(7), 1299–1300 (2012). Advanced issues in Artificial Intelligence and Pattern Recognition for Intelligent Surveillance System in Smart Home Environment
107. Rajpoot, Q.M., Jensen, C.D.: Video surveillance: privacy issues and legal compliance. In: Kumar, V., Svensson, J. (eds.), Promoting Social Change and Democracy through Information Technology. IGI global (2015)

108. Wang, X., Ellul, J., Azzopardi, G.: Elderly fall detection systems: a literature survey. Front. Robot. AI 7, 71 (2020)

109. Zhang, Z., Conly, C., Athitsos, V.: A survey on vision-based fall detection. In: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA ’15, New York, NY, USA. Association for Computing Machinery (2015)

110. Mshali, H., Lemlouma, T., Moloney, M., Magoni, D.: A survey on health monitoring systems for health smart homes. Int. J. Indus. Ergon

111. Bartula, M., Tigges, T., Muehlsteff, J.: Camera-based system for contactless monitoring of respiration. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2672–2675 (2013)

112. Walterscheid, I., Biallawons, O., Berens, P.: Contactless respiration and heartbeat monitoring of multiple people using a 2-d imaging radar. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3720–3725 (2019)

113. Karunaratne, I., Atukorale, A.S., Perera, H.: Surveillance of human-computer interactions: a way forward to detection of users’ psychological distress. In: 2011 IEEE Colloquium on Humanities, Science and Engineering, pp. 491–496 (2011)

114. Kang, M., Xia, L., Chen, H.: Research on the crowd abnormal behavior recognition in surveillance video based on modified social force model. In: 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), pp. 101–106 (2019)

115. Hossain, M.S., Muhammad, G., Guizani, N.: Explainable ai and mass surveillance system-based healthcare framework to combat covid-i9 like pandemics. IEEE Network 34(4), 126–132 (2020)

116. Al Hossain, F., Lover, A.A., Corey, G.A., Reich, N.G., Rahman, T.: Flusense: a contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4(1) (2020)

117. Villarroel, M., Chaichulee, S., Jorge, J., Davis, S., Green, G., Arteta, C., Zisserman, A., McCormick, K., Watkinson, P., Tarassenko, L.: Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. NPJ Digital Med. 2(1), 128 (2019)

118. Irtija, N., Sami, M., Ahad, M.A.R.: Fatigue Detection Using Facial Landmarks. In: 4th Int. Symposium on Affective Science and Engineering, and the 29th Modern Artificial Intelligence and Cognitive Science Conference (ISAIE-MAICS), WA, USA, 2018

119. Sikander, G., Anwar, S.: Driver fatigue detection systems: a review. IEEE Trans. Intell. Transp. Syst. 20(6), 2339–2352 (2019)

120. Guede-Fernández, F., Fernández-Chimeno, M., Ramos-Castro, J., García-González, M.A.: Driver drowsiness detection based on respiratory signal analysis. IEEE Access 7, 81826–81838 (2019)

121. Han, H., Jang, H., Yoon, S.W.: Driver head posture monitoring using mems magnetometer and neural network for long-distance driving fatigue analysis. In: 2019 IEEE SENSORS, pp. 1–4 (2019)

122. Mueid, R.M., Ahmed, C., Ahad, M.: Pedestrian activity classification using patterns of motion and histogram of oriented gradient. J. Multimodal User Interfaces 1–7 (2015). Springer

123. Gong, G., Wang, X., Mu, Y., Tian, Q.: Learning temporal co-attention models for unsupervised video action localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020

124. Cioppa, A., Deliege, A., Giancola, S., Ghanem, B., Droogenbroeck, M.V., Gade, R., Moeslund, T.B.: A context-aware loss function for action spotting in soccer videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020

125. Kozlowski, K., Korytkowski, M., Szajerme, D.: Visual Analysis of Computer Game Output Video Stream for Gameplay Metrics, pp. 538–552. Springer, Cham (2020)
126. Hussein, A., Gaber, M.M., Elyan, E., Jayne, C.: Imitation learning: a survey of learning methods. ACM Comput. Surv. 50, 21:1–21:35 (2017)
127. Torabi, F., Warnell, G., Stone, P.: Recent advances in imitation learning from observation. In: IJCAI (2019)
128. Atkeson, C., Schaal, S.: Robot learning from demonstration. In: ICML (1997)
129. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. Robot. Auton. Syst. 57, 469–483 (2009)
130. Jette, A.M.: The promise of assistive technology to enhance work participation (2017)
131. Ahmed, M., Idrees, M., ul Abideen, Z., Mumtaz, R., Khalique, S.: Deaf talk using 3D animated sign language: a sign language interpreter using Microsoft’s kinect v2. In: 2016 SAI Computing Conference (SAI), pp. 330–335 (2016)
132. Aloysius, N., Geetha, M.: Understanding vision-based continuous sign language recognition. Multimedia Tools Appl. 79(31), 22177–22209 (2020)
133. Nishimori, M., Saitoh, T., Konishi, R.: Voice controlled intelligent wheelchair. In: SICE Annual Conference, pp. 336–340 (2007)
134. Bai, J., Lian, S., Liu, Z., Wang, K., Liu, D.: Smart guiding glasses for visually impaired people in indoor environment. IEEE Trans. Consum. Electron. 63(3), 258–266 (2017)
135. Aafaq, N., Zulqarnain Gilani, S., Liu, W., Mian, A.: Video description. ACM Comput. Surv. (CSUR) 52, 1–37 (2020)
136. Hoffmann, F., Tyroller, M.I., Wende, F., Henze, N.: User-defined interaction for smart homes: voice, touch, or mid-air gestures? In: Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, MUM ’19, New York, NY, USA. Association for Computing Machinery (2019)
137. Rossi, M., D’Avenio, G., Morelli, S., Grigioni, M.: Augmented reality app to improve quality of life of people with cognitive and sensory disabilities. In: 2020 IEEE International Workshop on Metrology for Industry 4.0 IoT, pp. 59–62 (2020)
138. Kanno, K.M., Lamounier, E.A., Cardoso, A., Lopes, E.J., de Lima, G.F.M.: Augmented reality system for aiding mild alzheimer patients and caregivers. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 593–594 (2018)
139. Chu, F.J., Xu, R., Zhang, Z., Vela, P.A., Ghovanloo, M.: The helping hand: an assistive manipulation framework using augmented reality and tongue-drive interfaces. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2158–2161 (2018)
140. Masciò, S., Amedì, A.: Blind in a virtual world: mobility-training virtual reality games for users who are blind. In: 2015 IEEE Virtual Reality (VR), pp. 341–342 (2015)
141. Rashid, Z., Melià-Seguí, J., Pous, R., Peig, E.: Using augmented reality and internet of things to improve accessibility of people with motor disabilities in the context of smart cities. Future Gener. Comput. Syst. 76, 248–261 (2017)
142. Chien-Yu, L., Chao, J., Wei, H.: Augmented reality-based assistive technology for handicapped children. In: 2010 International Symposium on Computer, Communication, Control and Automation (3CA), vol. 1, pp. 61–64 (2010)
143. Gauci, J., Conti, E., Liang, Y., Virochsiri, K., He, Y., Kaden, Z., Narayanan, V., Ye, X., Chen, Z., Fujimoto, S.: Horizon: Facebook’s open source applied reinforcement learning platform. arXiv preprint arXiv:1811.00260 (2018)
144. Stephanidis, C., Salvendy, G., Antonia, M., Chen, J.Y., Dong, J., Duffy, V.G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L.P., Guo, Y.: Seven HCI grand challenges. Int. J. Human–Comput. Interact. 35(14), 1229–1269 (2019)