

LiDAR TECHNOLOGY FOR HUMAN ACTIVITY RECOGNITION: OUTLOOKS AND CHALLENGES

Omar Rinchi, Hakim Ghazzai, Ahmad Alsharoa, and Yehia Massoud

ABSTRACT

The smart and autonomous learning and recognition of human activities will certainly lead to an incredible progression for several applications and services in public healthcare, education, entertainment, safety and security, and more. With the recent advances in artificial intelligence, signal processing, and computational capabilities, light detection and ranging (LiDAR) technology can play an instrumental role to revamp current human activity recognition (HAR) systems. In this magazine, we investigate the potential of using LiDAR sensors as a novel technology enabling complex and real-time HAR applications. We first overview the HAR categories and the existing state-of-the-art technologies: video-based cameras, depth sensors, wearable devices, and WiFi. Afterward, we delve into the integration of LiDAR technology to perform HAR applications. Then, we discuss the advantages and drawbacks of utilizing LiDAR technology for HAR and compare it with the existing techniques. Finally, we discuss the challenges that need to be addressed to enable advanced LiDAR-based HAR applications.

INTRODUCTION

The deep monitoring and understanding of people's behaviors and activities have captivated increasing interest over recent years. This problem, known as human activity recognition (HAR), has attracted a lot of attention from the research and engineering communities to develop innovative solutions providing faster and more efficient preventive mechanisms.

We define the HAR as recognizing of a single human action or multiple humans actions using a series of observations and measurements of the humans' gestures, actions, behaviors, interactions, and the surrounding environment. A large number of applications fit under the scope of HAR including healthcare, mental health, and behavioral disorders, early childhood education, elderly nursing homes, schools, sports analytic, hazardous environments safety, and crowd management [1]. Such technologies, relying on various internet of things (IoT) sensors, are developed to prevent and mitigate incidents in different environments such as preventing elderly people or patients from falling, predicting students' illnesses in schools, avoiding crowd collisions and injuries, and improving the safety of workers in hazardous areas.

Recently, light detection and ranging (LiDAR) technology has been introduced as a novel technology that has the potential to significantly improve the capabilities of sensing systems. It utilizes eye-safe laser technology to calculate the distance between the sensor aperture and the object hit by the laser. By emitting multiple intense and focused laser beams, LiDAR sensor can create a three-dimensional (3D) digital map that measures angles and time differences between the emitted and reflected pulses and hence, provide a detailed representation of the surrounding environment [2]. Historically, LiDAR technology has been introduced to the public in 1971 during the Apollo 15 mission where an unsophisticated LiDAR system has been used to map the moon's surface. Thereupon, the use of LiDAR

sensors has been limited to a few applications in meteorology and astronomy. Just over the last decade, LiDAR technology has become widely used in many cutting-edge technologies. The most famous one is the automotive industry where LiDAR sensors are embedded in autonomous vehicles to detect surrounding obstacles and enable collision-free navigation. In physics and astronomy, many observatories are now supported with LiDAR sensors to perform accurate measurements with millimeter precision. Moreover, several military and civil applications (e.g., agriculture and transportation) are now using LiDAR sensors or LiDAR-equipped robots to perform complex sensing tasks.

With its demonstrated effectiveness in such critical applications and ability to map discrete changes in wide areas at very high resolutions and the possibility to manipulate its generated data with data analytic tools, LiDAR technology can play a significant role in real-time human and crowd behavior monitoring and recognition. In this magazine, we investigate the use of LiDAR sensors to monitor real-time human motions as a means to infer abnormal or unexpected motion behaviors in private and public environments. To the best of our knowledge, this magazine is the first to discuss the potential and challenges of a standalone LiDAR-based HAR solution. Its contributions can be summarized as follows:

1. Introduce the LiDAR technology as an effective mean to assess human activity with a detailed focus on the various possible modes of operation and outcome.
2. Investigate the advantages and limitations of utilizing LiDAR sensors with respect to other IoT sensors to perform real-time, remote, and confidential HAR.
3. Examine the state-of-art approaches and architectures, specifically the deep learning techniques, that can be utilized to conduct LiDAR-based HAR.
4. Discuss the open challenges related to the implementation of LiDAR technology for HAR applications.

HAR STATE-OF-THE-ART TECHNOLOGIES

Nowadays, there are several IoT technological solutions employed to identify and infer information about human behaviors and activities. The list includes video-based cameras, depth sensors, wireless signals, and sensor-based wearable devices. These approaches have shown promise in monitoring human

Omar Rinchi and Ahmad Alsharoa are with the Electrical and Computer Engineering Department, Missouri University of Science and Technology, Rolla, USA.

Hakim Ghazzai and Yehia Massoud are with the Innovative Technologies Laboratories (ITL), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

Digital Object Identifier: 10.1109/IOTM.001.2200199

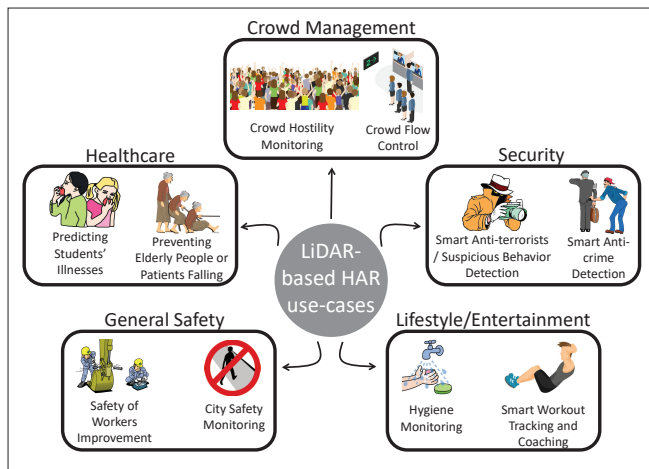


FIGURE 1. LiDAR-based HAR spans over a wide range of applications and use cases including healthcare, public safety & security, and crowd management.

and crowd behaviors; however, they are characterized by many drawbacks. In the sequel, we list the foremost state-of-the-art technologies used in HAR.

Video-based HAR: Video-based monitoring is the most commonly used method to provide insightful information about human behavior and activities, e.g., to monitor people in public areas and improve home and business security. Increased research efforts have been spent on fostering the use of cameras for HAR applications, especially with the emergence of machine learning and computer vision. Several innovative deep-learning algorithms have been developed to effectively recognize body motion and identify crowd behavior by analyzing snapshots or time series of images. Object detection, tracking the targets in real-time, or crowd counting were the main objectives of innovative approaches such as convolutional neural networks (CNN) (e.g., YoloV6) and recurrent neural networks (RNN) (e.g., long short-term memory (LSTM)) and their variations [1]. However, not all humans are willing to be monitored via traditional video or patient-engaged video surveillance due to factors such as threatened privacy and autonomy, even if monitoring might otherwise improve their well-being and safety. Furthermore, due to human privacy concerns, the use of cameras in private establishments and even in public places such as schools, hospitals, and shopping malls, is restricted by law in many countries. Camera technology can be negatively impacted by shadows, sunlight, and dim lighting, which can cause issues in night video monitoring for in-patient care and nursing homes. red-green-blue (RGB) images can be subject to optical illusions while robust background subtraction is also challenging for RGB-based cameras.

Depth Sensors: Depth sensors for HAR are range finder sensors that can produce 3D depth images. They are capable of capturing colors, textures, and depth resulting from RGB-depth (RGB-D) images. The existing depth sensors utilize one of two approaches to achieve this task:

- Time-of-flight (ToF) sensors that send light pulses from the sensor to the object and measure the distance using the received pulse ToF
- Infrared sensors that projects a pattern of infrared dots and uses an infrared camera to capture these patterns, measures the intensity of the infrared light, and hence, calculates the depth of the objects with respect to the sensor [3].

Both approaches overcome some of the limitations of conventional RGB cameras. On one hand, they are not color/light-dependent, which can solve any potential optical illusion. On the other hand, moving from two-dimensional (2D) to 3D can make processing easier as background subtraction can be done by specifying the region of interest (ROI) and analyzing 3D imag-

es. Nevertheless, the infrared light-based depth sensors can only work indoors due to their exposure to sunlight, which will degrade the performance of the detection. The ToF as an alternative can be subjected to the typical drawbacks of radio communication where multi-path propagation can significantly degrade the performance as the light pulses might be reflected from different surfaces, which may distort the computed distances.

Wearable Devices: Mobile and wearable devices, which are supported with built-in sensors such as accelerometers, gyroscopes, GPS, and medical sensors, provide non-visual information about their holders and are widely used to monitor the status of individuals, mainly for localization and healthcare purposes. Over the last decades, significant technological advances have been witnessed in this direction thanks to the emergence of wireless body area networks and smart electronic gadgets [4]. While these devices offer a layer of privacy protection that video cameras do not, they require users to remember to don/off the devices. For some humans, such as patients with cognitive or memory impairments, motion monitoring may be helpful and may facilitate a higher degree of autonomy, but these memory components represent a critical failure point. Additionally, they do not offer a robust monitoring overview for serial and contingent motion behaviors as they are usually designed to focus on specific aspects such as heart rate or blood glucose without providing any information about the actions and interactions undertaken by their holders.

WiFi: WiFi, a.k.a, IEEE 802.11 protocol, can be a potential solution for some HAR applications where multi-path propagation can be exploited to detect human activities and behaviors. The different movements and gestures of the human body can create a unique multi-path propagation pattern that can be detected by the receiver. Analyzing both the amplitude and phase of the channel state information (CSI) of the received signals can lead to a proper understanding of human gestures and movements [5]. With the recent advances in wireless technology especially with the reconfigurable intelligent surfaces (RIS), it is now possible to exploit the directed non-LoS signals to enable massive coverage of the targeted environments, which will enable massive coverage of the targeted environment. However, WiFi-based HAR suffers from some drawbacks such as being subject to different fading and shadowing effects. They are also weather (high-frequency), noise, and interference-dependent and can suffer from high attenuation losses due to long-range communication with high frequencies.

LIDAR TECHNOLOGY FOR HAR

In this section, we introduce LiDAR sensors as solutions for future HAR systems. We also discuss the currently utilized approaches in processing LiDAR data. LiDAR-based HAR can cover many applications in different domains as illustrated in Fig. 1.

LIDARS FOR HAR APPLICATIONS

In order to illustrate how LiDARs can fit into HAR scope, we start by categorizing LiDAR-based HAR applications based on the number of humans that the LiDAR scans and the HAR applications as described in Section III.C. These categories and how they will be processed are illustrated in Fig. 2.

Activity Recognition for Individuals: This category corresponds to the case of a single human located in a room with furniture in it. Humans can move freely inside the room and interact with the different objects in the surrounding. This scenario applies to many practical use cases such as in-patient care, transitional care, in-home care, nursing homes, and night monitoring. A robust set of physical body motions that may be indicative of concerning patient behavior (such as a fall, postural sway, unusual tremor, halted gate, twisted posturing, and so on.) needs to be defined for this purpose. Therefore, the activity recognition of individuals can be determined based on gestures, actions, behaviors, and interactions. Gestures are the movements of the body that can be used in nonverbal com-

munication. Actions are sequences of body movements such as walking and jumping. Behaviors are the analysis of certain actions for different contexts or locations, such as studying relaxation, stress, and productivity. Finally, interactions are when humans interact with their environment such as cooking (interacting with kitchen elements) or reading a book (interacting with the book).

Activity Recognition for Groups: This category encloses the scenarios of small groups of people sharing a common geographical area, usually closed areas, where the people can interact with each other and the objects in their surroundings. This category can correspond to the scenarios of nurseries and schools where children need to be remotely monitored to prevent incidents, or, for example, the case of police stations or detention centers to observe the behavior of detainees and criminals. Although it already involves the different classes that apply to individuals (gestures, actions, behaviors, and interactions), group activity recognition requires the consideration of an extra recognition layer focusing on the interactions between the co-located humans. The purpose is to detect abnormal behaviors for individuals or some members of the group, e.g., two children playing with each other. LiDAR for group activity recognition has the advantage of easier filtering. It can also fuse the HAR results of each individual in the group to come up with an upper-level conclusion about the activity of the group.

Crowd Activity Recognition: This category encloses the monitoring of large-scale events where large groups of people are gathering in the same area. Hence, the HAR in this scenario is focused on the behavior of the crowd itself as a single entity rather than the individuals as in the previous two categories. Such a scenario can be applicable in many practical use cases including crowd monitoring in public transport stations, sports events, and pilgrimages. Different crowd behavior patterns can be observed for this category by depicting the movement of the crowd. The patterns, flow, and density of the crowd can indicate whether an abnormal event is occurring or not. This can be helpful, especially to detect any dangerous behavior, e.g., panic or outbreak of violence.

LiDAR VS NON-LiDAR BASED HAR SENSORS

By scanning their surroundings, LiDAR sensors generate point clouds representing a set of data points that are hit by the laser. The point clouds can be processed with data analytic tools to automatically extract useful information enabling the inference of abnormal or unexpected behaviors for individuals, groups of individuals, or the crowd. The use of LiDAR technology presents extremely beneficial advantages for HAR applications. In fact, LiDAR technology can provide more depth information (e.g., objects' dimensions and distances between objects) without being subjected to any geometrical distortions, which is not the case with traditional and night vision cameras, wireless communication solutions, and wearable sensors. Although some types of cameras such as stereo cameras can provide highly detailed depth maps, such cameras have a limited range (i.e., a few meters) and the quality of the images is affected by dim lighting conditions. Moreover, LiDAR sensors are also immune to dark lighting conditions because they provide their own light source, making them well-suited for nighttime monitoring.

In Fig. 3, we compare LiDAR segmented pictures generated by our owned Livox Tele 15 sensor and images recorded by a regular camera in different scenarios. In Fig. 3a to 3d, it can be seen that LiDAR sensors are not affected by the light variation and they can work effectively in bright and dark environments. This is due to the consisting of active illumination units inside the sensors. Additionally, LiDAR technology provides extra valuable information that traditional video recording technologies do not. For instance, detecting fire smoke is possible by measuring the reflectivity indexes of the detected objects as illustrated in Fig. 3e and 3f.

Further, unlike traditional capturing technologies, LiDAR

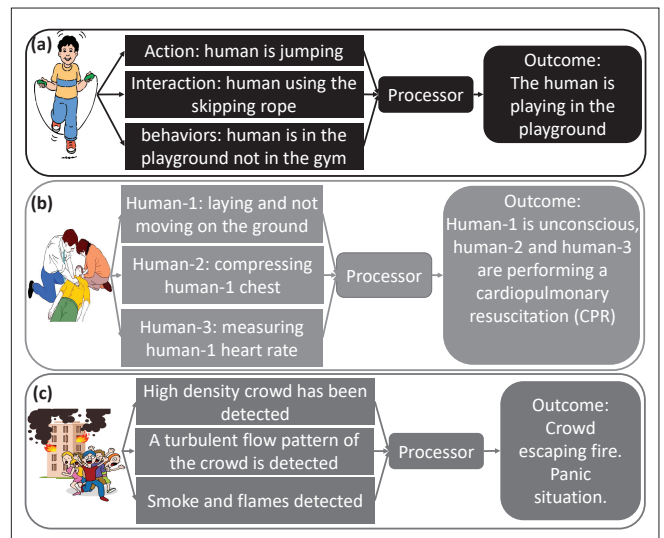


FIGURE 2. Smart LiDAR-based HAR mechanisms can be employed to draw conclusions regarding human behavior: (a) for individuals, (b) for groups of people, and (c) for the crowd.

technology does not collect traditional videos, which greatly reduces the collection of personal identifiable information (PII) such as recognizing facial features or reading textual data. This can significantly help in protecting the privacy of users and their belongings and hence, increase the possibility to use them in private locations to monitor people without violating their privacy such as schools, hospitals, and workplaces.

The video-HAR technology approach suffers from occlusion problems, LiDAR technology can tackle these problems more efficiently as it uses 3D point cloud data rather than the 2D RGB images/videos utilized by video cameras. Different filtering techniques such as ROI-based filtering, and density-based clustering can exploit the characteristics of the 3D point cloud data leading to more efficient and easier processing with fewer possible occlusion challenges. On the other hand, LiDAR allows for even easier processing of the data that belongs to crowds with fewer possible occlusion problems as LiDAR utilizes point cloud data. Crowd activity can be recognized by analyzing the overall trend of the whole points in the cloud. More specifically, the whole set of points in the cloud are considered as one big object with no required clustering or object identification for any subset of points in the big cloud. This property of point cloud data allows crowd activity recognition with no occlusion challenges which cannot be performed with the 2D RGB images of video cameras. In Table 1, we provide a qualitative analysis comparing the utilization of LiDAR technology for HAR applications with existing techniques.

DEEP LEARNING ARCHITECTURES FOR LiDAR-BASED HAR IMPLEMENTATION

Studying HAR using LiDAR technology is still at the early stage and very few studies dealing with the problem exist. For example, in [6], the authors gathered 2D LiDAR data to perform human and non-human classifications. Then, a Kalman filter is employed to track the trajectories of humans. Similarly, in [7], the authors advocated the utilization of a 2D LiDAR sensor to detect human falls. The proposed method locates the individual position in the testing room and detects four classes of human activities: walking, standing, sitting, and falling. Unlike 3D LiDAR, the processing of 2D LiDAR data is quite similar to traditional image processing. Indeed, the authors employed the CNN, ResNet34, combined with the RNN, LSTM, to perform the classification on the 2D LiDAR frames. In [8], the authors employed both RGB camera images with 3D LiDAR data to detect human activity. The 2D images are utilized as ground truth to identify 2D ROI in the envi-

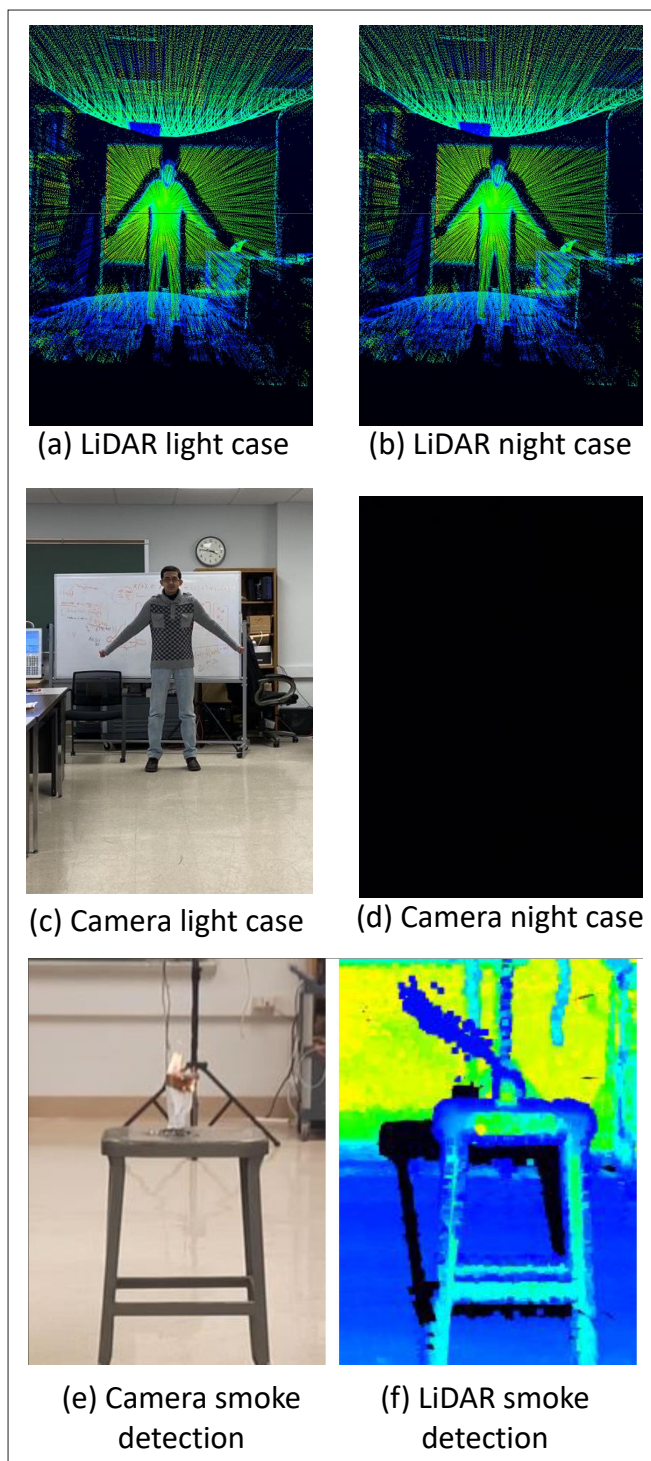


FIGURE 3. Testing using Livox Mid 40 LiDAR sensors.

ronment and a CNN is used to detect the activity using a 2D Fisher vector representation of the 3D LiDAR data. The method detects up to nine actions with over 90% accuracy. The actions the authors investigated were quite simplistic such as carrying a box, pushing a board, running, walking, and texting. LiDAR technology has been employed for crowd management purposes, especially for traffic monitoring. Usually, the target objects are vehicles or pedestrians. In most cases, LiDAR data is fused with other types of data, mainly video frames, to detect and track objects. One recent study has presented the architecture of an IoT-enabled 3D visualization system for autonomous crowd management suggesting the deployment of LiDAR sensors as well as other sensors such as gas sensors and regular and infrared camer-

as to collect multi-modal data. Artificial intelligence algorithms are proposed to be used to compute the locations of moving objects and detect fights, stampedes, fires, and other abnormal events [9].

LiDAR-based HAR consists of three major steps: pre-processing the raw data, applying the learning model, and post-processing the output of the learning model to obtain the final HAR outcome. These three steps are summarized in Fig. 4. The raw data is mainly LiDAR's range measurements across the three axes. In some scenarios, the reflectivity measurements are also included as extra raw data, which can help identify the nature of the materials with which the object hit by the laser are made. Pre-processing is the first step that converts the raw data into a suitable format for the learning model. Different pre-processing techniques can be performed, for instance, data filtering can remove the undesired points in the cloud. This can be mainly done by specifying a ROI, or by applying clustering techniques such as density-based spatial clustering of applications with noise (DBSCAN). Data rotating is rotating the data direction by multiplying with a rotation matrix (RM). Data labeling is assigning a label for each instance of data to be able to train the data using supervised learning techniques. Data augmentation can be also performed to increase the number of data that are associated with a certain label or to reduce the negative effects of sparsity in the data. The next step is applying a learning model that is usually in the form of a shallow neural network or a deep learning model. Due to the unique characteristics of LiDAR cloud data such as

- Sparsity
- Distribution in a varying density
- Unorderedness and more, traditional models such as CNNs fail to directly accomplish the learning task.

As a result, novel approaches have been proposed in the literature specifically to address point cloud-related challenges. They can be classified into three categories:

- Point-based approach: the model is trained using the raw measured point cloud. The PointNet as a popular example of this category is inspired by the spatial transformation networks (STN) in which the data is pre-processed such that the data is invariant to certain geometric transformations (i.e., rotations and translations) [10]. In addition, max pooling is used to create a global feature vector to solve the challenges that could result due to the unorderedness in the input point cloud. Finally, the PointNet utilizes multi-layer perception (MLP) for both classification and segmentation.
- Voxel-based approach: the model is trained using newly constructed four-dimensional (4D) tensors rather than the raw point cloud. The VoxelNet is a popular example in this category [11]. In VoxelNet, the environment is divided into multiple 3D cells (voxels) with fixed dimensions. The input point cloud is grouped into these voxels, randomly down-sampled to reduce complexity, and then encoded using stacked voxel feature layers to add features (i.e., the fourth dimension) for each 3D voxel. CNN can be utilized to train the model while region proposal networks are utilized for object detection.
- Hybrid approach: combinations of point-based approaches and voxel-based approaches are utilized. The PointVoxel CNN (PV-RCNN) is an example of this category. This method is a combination of the CNN based on voxels and PointNet set abstraction based on points. The framework starts with 3D voxel CNN to encode the point cloud into 3D feature volumes and generate 3D proposals. Afterward, it uses the voxel-to-keypoint scene encoding to convert the feature volumes voxels into feature keypoints. Finally, point-to-grid ROI feature abstraction is performed to transform the keypoint features to ROI grids for confidence prediction and proposal refinement.
- Other approaches that could not fit into any of the previous three categories include prediction networks such as the RNN with all its variations, SVM for classification, and complex-YOLO as a modification on YOLO for point cloud.

Post-processing is an optional step and can be case-depen-

	Video Based Camera	Depth Sensors	Wearables	WiFi	LiDAR
Applications	Most HAR applications except for medical and fitness applications	Most HAR applications except for medical and fitness applications	Mostly for navigation, tracking, medical, and fitness applications	Mostly for gestures detection, sensing, tracking, and localization	Most HAR applications except for medical and fitness applications
Data Output	RGB Images/Videos	RGB-D images/videos	Numerical data	Numerical data	2D/3D point cloud
Data Processing and Complexity Level	High	High	Low	Medium	High
Illumination Requirement	Required unless for thermal cameras	Required	Not required	Not required	Not required
LoS dependency	LOS dependent	LOS dependent	N/A	LOS independent	LOS dependent
Optical Illusions	Subject to	Impregnable	Impregnable	Impregnable	Impregnable
2D vs 3D	2D	3D	N/A	N/A	3D
Weather Condition Effect	Medium effect	Medium to high effect	No effect	Medium to high effect	Medium to high effect
Privacy	Reveals PII (including low resolution cameras)	Most depth sensors reveals PII	Save PII	Save PII	Save PII
Cost	Low to high	Low to high	Low to high	Medium to high	Low to high
Technology Maturity	High	Medium	High	High	Low
Other Remarks	N/A	• Usually indoor environments • Short distances	• Require to be wearable all the time • Limited to HAR for individuals	• Usually indoor environments • Require special equipment	• Can detect gases • Provide information about the object reflectivity

TABLE 1. A comparison between IoT technological solutions for human activity recognition.

dent. Tracking algorithms such as Kalman filter or sensor fusion techniques might be used in this step. It can also include processes for data analysis improvement, judgment, and visualization. The final HAR outcome of the LiDAR-based HAR can be in different forms such as segmentation, activity recognition, trajectory identification, and tracking.

LIDAR-BASED HAR RESEARCH CHALLENGES

The research trend is to design deep learning architectures and algorithms operating on the LiDAR 3D point cloud data to extract insightful information about human activity for different HAR applications. Novel strategies for detection, classification, and tracking of body movements, human interactions with their peers as well as their surroundings, and crowd motion patterns need to be developed. This must be performed while considering the challenges related to the large volume of LiDAR data that might be collected. In the following, we delve into the main research challenges associated with the utilization of LiDAR sensors for HAR applications.

LIDAR DATA ASSESSMENT AND COLLECTION

One of the main challenges related to the implementation of artificial intelligence on LiDAR data for HAR applications is the absence of LiDAR datasets, especially the annotated ones that can be employed to effectively train the machine and deep learning models. In fact, even the existing studies discussed previously, are mainly dealing with 2D LiDAR data. Creating annotated and labeled datasets of 3D point clouds depicting the different HAR scenarios is a challenging task. The goal is to identify sets of normal and abnormal activities for the different scenarios related to body gestures, actions, interactions, and crowd patterns. Unlike the few previous studies that are focusing on straightforward classes such as walking, running, sitting, laying down, or any similar, more complicated classes detailing the behaviors of individuals and the crowd should be identified, potentially in a separate

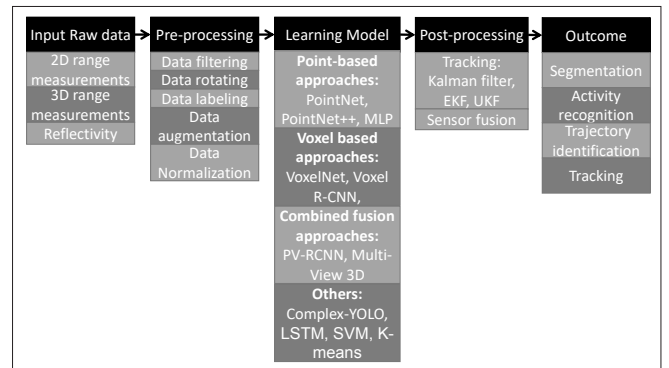


FIGURE 4. Block diagram summarizing the different processes that can be involved in the general steps of LiDAR-based HAR.

manner for each HAR use-case and application. Moreover, the dataset collection should consider the different types of LiDAR sensors existing in the market as these sensors are characterized by different specifications such as the resolution, the range, the reflectivity levels, and the field of view angles.

All these specifications have a direct impact on the size of the collected data and the efficiency of the learning models that will be utilized. Therefore, it is vital to pinpoint the important parameters of the LiDAR sensors that can influence the HAR problem. Since the ultimate goal of the HAR problem is to achieve a tradeoff between deep learning model efficiency, computational speed, and minimum volume of LiDAR data, several types of LiDARs such as short- and long-range LiDAR sensors should be experimentally tested for such applications.

One approach to generate and analyze LiDAR's data is to use computer-programmed software to generate virtual LiDAR 3D cloud data. For instance, in Fig. 5, we use the Webot soft-

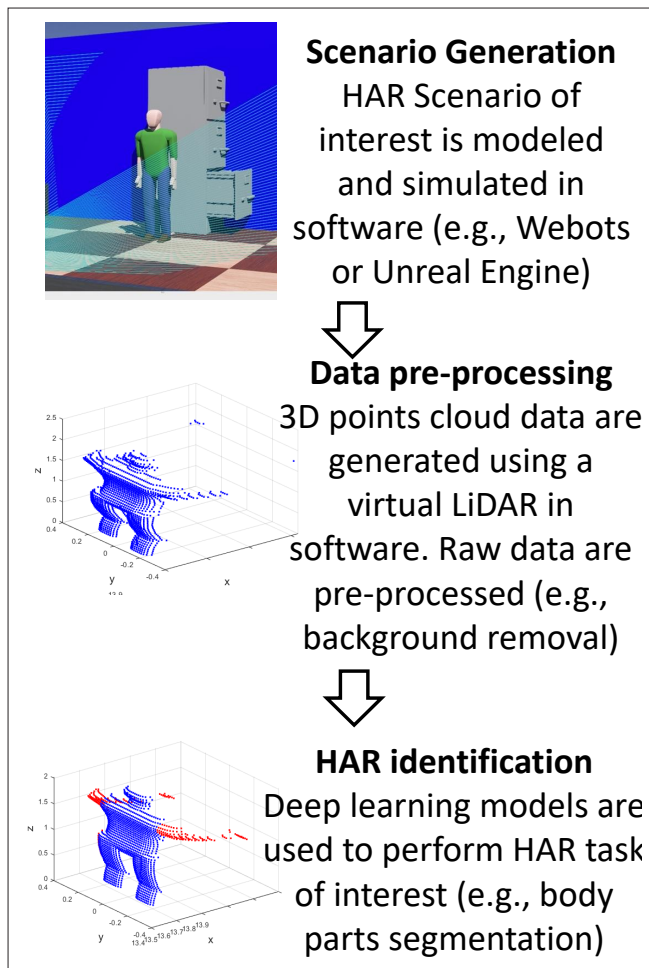


FIGURE 5. Example of 3D LiDAR point cloud data generation and analysis. The human behavior is simulated on Webot, a computer-programmed software, to generate a 3D point cloud. After data pre-processing and filtering, the human gesture is identified using deep learning algorithms. The virtual LiDAR utilizes 53 scanning layers, and the virtual human is placed around 15 m away from the virtual LiDAR.

ware to generate virtual 3D cloud data to simulate realistic situations. These virtual data is used to train a MLP to identify the gestures of virtual human. This approach can easily reduce time, effort, and costs; it can easily generate millions of LiDAR measurements in a very short time. On the other hand, it is possible to generate data that belongs to complex and risky scenarios that are hard to obtain using a normal setup. Using such approaches, the virtual software setup will help annotate data and train, complex models. The final neural network models can be either directly utilized or applied after fine-tuning and transfer learning to real-world situations. In addition to that, utilizing digital twins and generative adversarial networks (GANs) can also solve the limitations of LiDAR data. Digital twins can create a digital replica of the environment of interest, allowing experiments to be performed on the digital model instead of the physical model. On the other hand, GANs can generate new and fake data sets, including human-shaped point cloud data, which can support the efforts toward LiDAR data analysis [12].

AUTONOMOUS LEARNING MODELS FOR ACTIVITY RECOGNITION

Pre-processing and analyzing the sequences of LiDAR data to enable the detection, classification, and tracking of the behaviors of individuals and the crowd using machine and deep learning algorithms are a second challenge that needs to be taken into account in LiDAR-based HAR applications. Compared to video

cameras, LiDAR sensors may generate a huge volume of data that needs to be processed. Therefore, a multi-stage filtering technique is recommended to pre-process the data and eliminate redundancy. The technique needs to specify the ROI of the human, group of humans or the crowd, and eliminate unwanted data, e.g., corresponding to ground/ceiling/background, to speed up the subsequent data processing steps and training of the learning models. Afterward, a novel non-uniform voxel grid can be employed when needed to produce a better sampling such that more points are deleted from denser objects compared to less-dense objects when needed. Finally, a density-based clustering algorithm can be employed to cluster the objects.

Several learning models processing LiDAR 3D point cloud data applied to different applications exist. These techniques could be utilized in the context of HAR. For instance, PointNet++ has attracted significant interest over the last few years in smart mobility due to its ability to process the 3D cloud data points directly with high accuracy. For instance, it is employed with autonomous vehicles for pedestrian monitoring services. On the other hand, voxel-based detection can achieve a reliable computational complexity. Support vector machines, sensor fusion techniques, and neural networks can also be applied to LiDAR data as well. Therefore, existing learning approaches could be upgraded and accommodated to the context of HAR by applying transfer learning and tuning to improve the models' accuracy in detecting real human and crowd behaviors. Still, novel approaches to enable the detection, classification, and tracking of targets in HAR applications are required with the challenge of performing real-time activity recognition with minimum data information and delays. Among the potential challenges is to understand the interactions of the human with the surrounding objects, track the motion of its body parts (arms, legs, and head), classify the behavior of the crowd, track its flow, and count the number of people in crowd scenes.

MULTI-LIDAR INTEGRATION

The deployment of multi-LiDAR systems for HAR applications is very useful in many situations, especially for group and crowd management applications. Multiple LiDAR sensors can be installed in different locations to perform automatic and continuous sensing tasks to collect the data in the ROI. This allows the coverage of wide areas, provides different observation angles and helps cope with the absence of direct LoS links between a single LiDAR sensor and the targets. Different types of LiDAR sensors can be used for such tasks resulting in collecting heterogeneous 3D point cloud data. Exploring the joint combination of the collected data and its use to estimate the current state of the target, i.e., individual, groups, or crowd, while devising high-precision data fusion techniques adequate for the HAR scenarios are the main challenges for the multi-LiDAR systems.

The collected data must be synchronized and concurrently exploited. Hence, artificial intelligence techniques for 3D point cloud data fusion are required to overcome these challenges for fast decision-making. According to the level of LiDAR information fusion representation, the level of fusion can be divided into three levels:

1. Point cloud-level fusion: directly fuse the raw data collected by the LiDARs,
2. Feature-level fusion: first perform feature extraction (such as extracting shape features or motion features), and then fuse these features in the fusion center to obtain the fused comprehensive features
3. Decision-level fusion: extract features independently for each LiDAR to generate feature vectors, make decisions through the feature vectors, and then synthesize the decisions between different sensors according to a certain algorithm to obtain a final decision.

Finally, another challenge related to the employment of multiple LiDAR sensors is the LiDAR infrastructure-planning problem. This problem can be seen as an optimization program

that might be optimally solved. The goal is to determine an optimized placement and orientation of the LiDAR sensors in the region of interest as well as their specifications that can guarantee an effective operation of the system without causing exhaustive network congestion and complex data processing. Greedy and meta-heuristic algorithms can be investigated to achieve low-complexity solutions for this problem.

COMPUTATION AND DATA MANAGEMENT OF LiDAR DATA

LiDAR sensors produce huge amounts of data when scanning their surrounding areas, and the amount of data increases with their precision and resolution. Considering that, the aim of the LiDAR systems for HAR is to reduce the local computational burden on the local servers, integrating the mobile edge computing (MEC) technology in the design of the LiDAR infrastructure would be appropriate to perform local decision-making and reducing the burden on the cloud. Not only in terms of data processing but also to combat network congestion [13]. Indeed, the main bottleneck of the cloud-only solution is the data transfer and routing over the networks, which usually suffer from heavy congestion with very large volumes of data transmission.

MEC task offloading is an active area of research that aims to overcome the network bottleneck in the cloud by incorporating edge computing in the backhaul. Computers “on the edge” of the network are lower-power devices that exist far outside of the data center, and much closer to the network gateway that LiDAR can communicate through. Although MEC devices are not as powerful as the cloud servers are, they are much closer to the data source, so the network transmission latency is far lower. This reduction in network latency offsets the slower processing speed, especially as there are large performance boosts in MEC processors over each generation. Hence, a tradeoff between processing speed and network overload could be determined to enable effective LiDAR operation for real-time HAR applications.

The MEC architecture offers multiple new destinations and routes that communications traffic may be oriented to. This requires novel backhaul routing and task allocation solutions to ensure efficient utilization of the computational resources while taking into account LiDAR data specifications and the quality of service requirements of HAR applications.

Another way to cope with the huge LiDAR data generation is to develop data management such as dimension-reduction techniques. These techniques can be divided into two categories: the first category aims to trim the extra data and keep only the most impactful features such as backward elimination, forward elimination, and random forests. The second category aims to compute a new small combination of the old large features such as the principle component analysis (PCA), the factor analysis (FA), Kernel PCA, and multidimensional scaling (MDS). Additionally, sparse recovery algorithms can exploit the sparse nature of the point cloud to compress the data into smaller dimensions while recovering the full size later when needed.

LiDAR IN MULTI-MODAL HAR SYSTEMS

LiDARs may not provide all the necessary information for some HAR applications, such as medical-related HAR, requiring biometrics. Sensor fusion techniques can improve system reliability and robustness by combining LiDAR with other sensors such as WiFi, wearable devices, or others.

Similar to multi-LiDAR fusion, LiDAR and other sensors fusion can be performed in three levels: decision level, feature level, and raw-data level. At the decision level, the results are fused at a high level of abstraction. The data from each sensor is given to an independent algorithm where detection, segmentation, or classification is performed. Afterward, the outputs of each algorithm are merged using different strategies like Bayesian-based fusion, or voting-based fusion. The main advantage of this type of fusion is that it can allow the use of

incompatible data coming from different sensors and remove data registration and data alignment errors. On the other hand, the feature level is where the features are combined after feature extraction, and the raw data level is where the point cloud are decorated with the corresponding other sensor raw-data at a lower level. These approaches provide higher accuracy compared to the decision level while both are challenged in how to integrate the raw data of the LiDAR with the one of the other sensors. Most of the existing works in the literature such as the TransFusion and the DeepFusion target the LiDAR/Camera fusion for object detection [14]. Other opportunities for more complex HAR applications involving different LiDAR/IoT sensor fusion are worth for further investigation.

ENVIRONMENTAL EFFECT MITIGATION

LiDARs face environmental challenges outdoors, such as rain, snow, and fog, which can degrade their performance by reflecting, scattering, or absorbing the transmitted beams. This can cause false positive and false negative detections and decrease the maximum detection range. The authors in [15] investigated the performance of the LiDAR under the rain. The authors have utilized Unreal Engine to simulate how LiDAR will work under different rain intensities. Their results also show that LiDAR starts to lose more points (i.e., false negative detection) with increasing distance under the rain.

Different procedures can be used to encounter environmental effects, one way is to handle this problem in software. In this case, environmental effects are considered as noise and the algorithms can be designed to reduce the noise effects. Improving the hardware can also improve the performance, for instance, different wavelengths of LiDAR react differently to environmental noise. On the other hand, LiDARs using frequency-modulated continuous wave (FMCW) perform better under bad weather conditions compared to ToF-based LiDARs. Sensor fusion is an alternative solution to improve the performance as LiDAR can be integrated with other sensors to mitigate the environmental effects.

CONCLUSION

In this magazine, we have investigated LiDAR technology as a promising HAR enabler for individuals, groups, and crowds. We examined the possible capabilities of LiDAR-based HAR and compare them with the current existing HAR approaches. Afterward, we discussed some limitations of the technology and spotted the light on the challenges that it can face. We have shown that this technology for HAR applications is still at an early development stage and a huge research effort is required to enable real-time LiDAR-based HAR operation.

REFERENCES

- [1] P. Pareek and A. Thakkar, “A Survey on Video-Based Human Action Recognition: recent Updates, Datasets, Challenges, and Applications,” *Artif. Intell. Rev.*, vol. 54, no. 3, Mar. 2021, pp. 2259–2322.
- [2] S. Chen et al., “3D Point Cloud Processing and Learning for Autonomous Driving: Impacting Map Creation, Localization, and Perception,” *IEEE Signal Proc. Mag.*, vol. 38, no. 1, Jan. 2021, pp. 68–86.
- [3] H. Raelis, M. Kazemi, and S. Shirmohammadi, “Human Activity Recognition with Device-Free Sensors for Well-Being Assessment in Smart Homes,” *IEEE Instrum. Meas. Mag.*, vol. 24, no. 6, pp. 46–57, Aug. 2021.
- [4] E. Ramanujam, T. Perumal, and S. Padmavathi, “Human Activity Recognition With Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review,” *IEEE Sens. J.*, vol. 21, no. 12, June 2021, pp. 13,029–40.
- [5] H. Yan et al., “WiAct: A Passive WiFi-Based Human Activity Recognition System,” *IEEE Sens. J.*, vol. 20, no. 1, Jan. 2020, pp. 296–305.
- [6] F. Luo, S. Poslad, and E. Bodanese, “Temporal Convolutional Networks for Multiperson Activity Recognition Using a 2-D LiDAR,” *IEEE Internet Things J.*, vol. 7, no. 8, Aug. 2020, pp. 7432–42.
- [7] M. Bouazizi, C. Ye, and T. Ohtsuki, “2D LiDAR-Based Approach for Activity Identification and Fall Detection,” *IEEE Internet Things J.*, pp. 1–1, Nov. 2021.
- [8] J. Roche et al., “A Multimodal Data Processing System for LiDAR-Based Human Activity Recognition,” *IEEE Trans. Cybern.*, June 2021, pp. 1–14.
- [9] Q. Yu et al., “Intelligent Visual-IoT-Enabled Real-Time 3D Visualization for Autonomous Crowd Management,” *IEEE Wireless Commun.*, vol. 28, no. 4, Aug. 2021, pp. 34–41.
- [10] C. R. Qi et al., “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation,” *Proc. 16th IEEE Conf. Computer Vision and Pattern Rec-*

ognition (CVPR'17), Minneapolis, MN, USA, July 2017. 9

- [11] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection," *Proc. 17th IEEE Conf. Computer Vision and Pattern Recognition (CVPR'18)*, Salt Lake City, UT, USA, Dec. 2018, pp. 4490–99.
- [12] R. Li et al., "PU-GAN: A Point Cloud Upsampling Adversarial Network," in *Proc. 17th IEEE/CVF Int'l. Conf. Computer Vision (ICCV'19)*, Seoul, Korea (South), Oct. 2019.
- [13] M. C. Lucic et al., "A Latency-Aware Task Offloading in Mobile Edge Computing Network for Distributed Elevated LiDAR," *Proc. 47th IEEE Int'l. Symp. Circuits and Systems (ISCAS'20)*, Seville, Spain, Sept. 2020, pp. 1–5.
- [14] X. Bai et al., "Transfusion: Robust LiDAR-camera fusion for 3D object detection with transformers," *Proc. 21th IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR'22)*, New Orleans, LA, USA, Sept. 2022, pp. 1090–99.
- [15] J. P. Espineira et al., "Realistic LiDAR with Noise Model for Real-Time Testing of Automated Vehicles in A Virtual Environment," *IEEE Sensors J.*, vol. 21, no. 8, Feb. 2021, pp. 9919–26.

BIOGRAPHIES

OMAR RINCHI [StM] (omar.rinchi@mst.edu) received a B.Sc. degree in Mechatronics Engineering from the University of Jordan, Amman, Jordan, in 2020. He is now a Ph.D. student with the Electrical and Computer Engineering Department at Missouri University of Science and Technology, Rolla, MO, USA.

HAKIM GHAZZAI [SM] (hakim.ghazzai@kaust.edu.sa) received the Ph.D. degree in electrical engineering from King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in 2015. He is currently a Research Scientist at KAUST. His general research interests include wireless networks, UAVs, Internet-

of-Things, and intelligent transportation systems. Since 2019, he has been on the Editorial Board of the *IEEE Communications Letters* and the *IEEE Open Journal of the Communications Society*.

AHMAD ALSHAROA [SM] (aalsharoa@mst.edu) was born in Irbid, Jordan. He received the Ph.D. degree from Iowa State University, USA, in May 2017. He is currently an Assistant Professor with the Electrical and Computer Engineering Department, Missouri University of Science and Technology (Missouri S&T). His current research interests include LiDAR detection and tracking, human activity recognition, smart systems, wireless networks, optical communications, high–low altitude platform communications, Internet of Things (IoT), energy harvesting, green wireless networks, and reconfigurable intelligent surfaces. Since 2021, he has been on the Editorial Board of the *IEEE Wireless Communications Letters* and the *Frontiers in Communications and Networks* journal.

YEHIA MASSOUD [F] (yehia.massoud@kaust.edu.sa) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, USA. He is currently a Professor and the Director of the Innovative Technologies Laboratories (ITL) at King Abdullah University of Science and Technology (KAUST), Saudi Arabia. His research focuses on the design of state-of-the-art innovative technological solutions that span over the broad range of technical areas including smart cities, autonomy, smart health, embedded systems, nanophotonics, and spintronics. He has authored over 350 articles in peer-reviewed journals and conferences. He was a recipient of the Rising Star of Texas Medal, the National Science Foundation CAREER Award, the DAC Fellowship, the Synopsys Special Recognition Engineering Award, and two best paper awards.