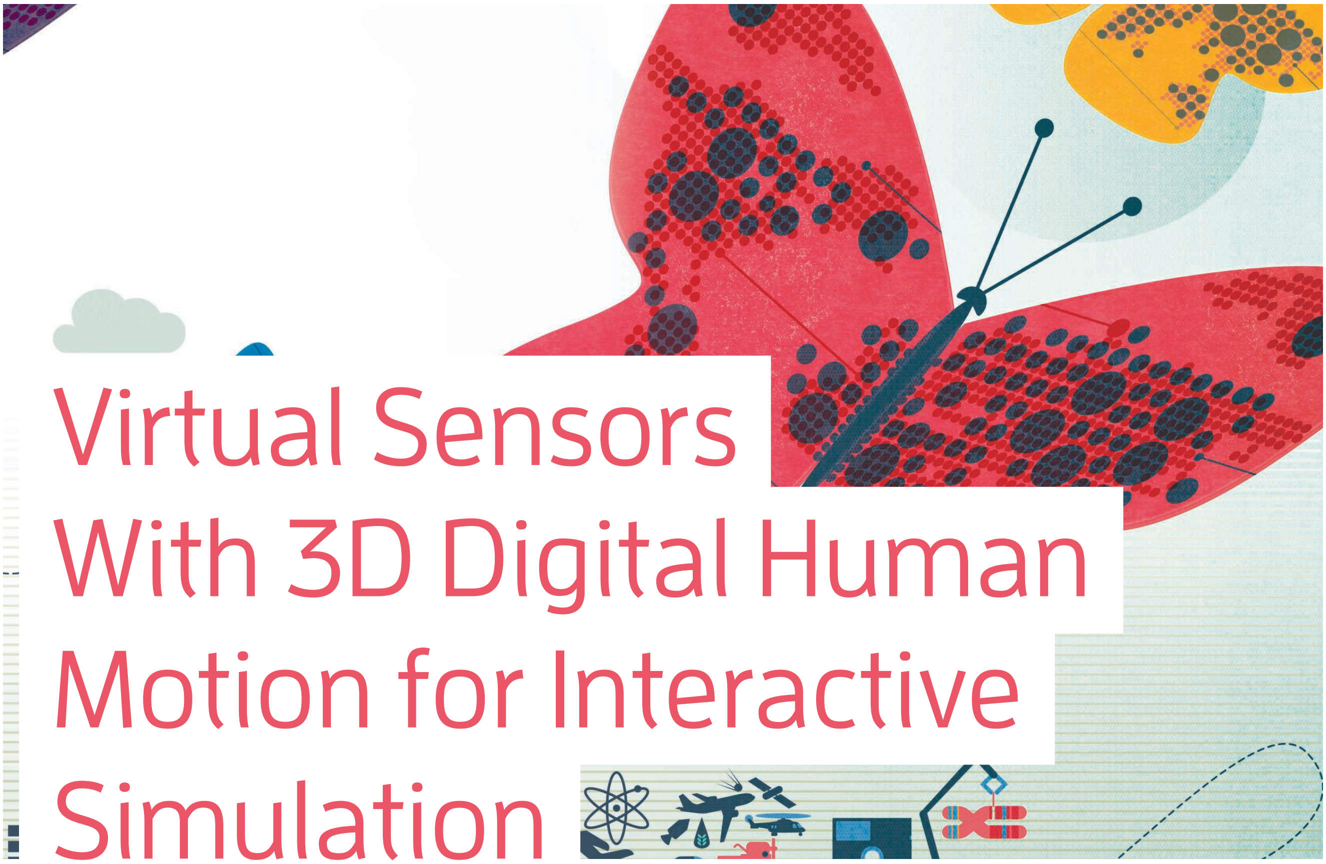

Virtual Sensors With 3D Digital Human Motion for Interactive Simulation

Chengshuo Xia, Xidian University
Yuta Sugiura, Keio University

This article presents a design of "virtual sensors" to collect low-dimensional sensor data from 3D digital human motion and create real-world applications through the "interactive simulation." It provides an opportunity to decrease dependency on real-world data requirements and gives more flexibility in the corresponding human activity-related applications.

Using sensors and machine learning (ML) to recognize human motion and build the activity-related applications has been successfully applied to assist people's daily lives, such as gesture interaction, motion tracking, and exergames.[1,2] As in Zhu et al.,[2] multimodal sensors have been used in various types of activity recognition systems. However, such systems have shown bottlenecks; that is, real sensor datasets are always required as prerequisites to train classifiers. As a result, the developed applications are mainly limited by the dataset's characteristics, including data modalities, sensor characteristics, and data types.

INTRODUCTION

Such issues affect the flexibility of users to employ the system. For example, when using a wearable activity recognition system, the specified sensor-wearing position cannot

Digital Object Identifier 10.1109/MC.2023.3239344
Date of current version: 13 November 2023

 0018-9162/23©2023IEEE

fit different users' body conditions and preferences. For a fitness-tracking system, users normally require various types of activity to be recorded based on their needs. However, these issues are determined by the needs of the data during the system design process. Once the system's characteristics need to be changed (for example, where the sensor is to be placed and the type of activities to be recognized), the data need to be recollected, and the systems must be redesigned.

To solve these problems, there have already been some efforts to generate synthetic sensor data to reduce the cost of real sensor data. Traditional synthetic sensor data generation typically relies on generative adversarial network (GAN) models.[3] Synthetic data can be generated by using real data to train the discriminator and generator, effectively augmenting datasets to train classifiers with higher robustness. For low-dimensional time series sensor data [for example, wearable inertial measurement unit (IMU) sensor data], several works have proposed to produce data. Sensegen[4] introduced a long short-term memory network as the generator to produce synthetic linear acceleration data. SensoryGAN[5] designed an unabridged model to generate the acceleration data by a 1D-convolutional neural network (CNN) model. ActivityGAN[6] employed 1D- and 2D-CNN structures, which showed better discriminator performance. Pham et al.[3] proposed a conditional GAN model to generate a synthetic walking step length by a waist-mounted IMU sensor. However, this approach still requires a large amount of real data to train the model and does not fundamentally improve the need to obtain real data at a high cost. In addition, another approach to generate synthetic sensing data by videos has emerged in recent years, which uses 2D characters in videos to infer 3D skeletal point displacement information and thus computes acceleration information for a specific body location, as in IMUTube.[7] Vid2Doppler[8] extracted the mesh of the subject in the video to calculate the synthetic Doppler radar signal and used it to augment the existing radar dataset. Although this approach can generate synthetic data without relying on existing real datasets, the generation process still lacks intuition. The generated low-dimensional sensor data are abstract, and the application scenario is limited to the augmentation of existing datasets. Therefore, designing, developing, and generating synthetic sensor data more intuitively and conveniently, can solve the costly dataset problem and create more application scenarios.

Traditionally, as an alternative to real people, digital humans have been used to evaluate ergonomics in virtual environments and to reduce the task of product development by simulating the posture and motion of real-world people.[9] Moreover, digital humans, combined with motion capture (MoCap) systems, can reproduce the motions of real people. Taking advantage of this, we introduced a new application of the 3D digital human that relies on the simulation of a physics engine capable of measuring changes in the virtual physical variables generated by the motion of the 3D digital human. Measuring the generated virtual physical variables to obtain the synthetic sensor data are called *virtual sensor detection*. The data detected by virtual sensors can help improve the development of relevant systems in the real world through "interactive simulation."

By reconstructing the motion of a 3D digital human, virtual sensors are designed in the virtual environment to detect the signal changes brought about by the motion of the digital human. These signals are used to replace real sensor signals in the traditional development process. Due to the virtualization of sensor data collection, this method will shorten the development process and reduce costs. More importantly, the low-cost data can bring more system flexibility and greatly enhance the convenience of end users. This article presents the development of an interactive simulation system based on this method, which allows developers to effectively check the characteristics of the developed system and adjust it to the actual situation while obtaining the required sensor data at a low cost. The following sections will introduce the implementation of several key techniques, including 3D human motion reconstruction and virtual sensor design, the applications with interactive simulation, future work, and challenges.

REPRODUCE HUMAN MOTION BY DIGITAL HUMANS IN A VIRTUAL ENVIRONMENT

Synthetic sensor data are obtained similarly using virtual sensors as sensor signals associated with human motion are obtained in the real world. After reconstructing human motion, the variables generated by human motion are measured in the same way as in the real world (for example, by placing the sensor in space or wearing it on the body). Thus, obtaining a 3D human motion sequence is a prerequisite for obtaining synthetic sensor data using virtual sensors. This section described how to get a 3D digital human motion sequence in a game engine.

Environment: Game engine platform

To support the reconstruction of digital human motion and producing the related sensor data, the high-performance and high-accessibility platform is considered. The Unity 3D game engine not only fuses the physics engine but is also popular in the gaming design field, which makes it possible to create various applications and can be widely accessed by end-users. Thus, the whole design was developed based on Unity 3D.[10]

3D digital human avatar

To simulate the real motions conducted by humans, creating the 3D digital human avatar, as the subject of movement execution in the virtual environment, is the first step. Generally, the hierarchical skeletal structure organizes the humanoid avatar model, and the root node is normally selected as the center of gravity of a humanoid model. The other body parts are connected following the parent-to-child structure (Figure 1).

To get a 3D avatar, there are already many resources for accessing 3D humanoid models, such as Mixamo.[11] The resources have been embedded by a limited predefined motion and some various models with a different appearance and are typically applied to the development of 3D games to increase the reality. In addition to existing animated character models, other types of software can help develop a more refined model, such as the exclusive skinned multiperson linear model from MakeHuman.[12] Monocular video-based 3D character model extraction has also received wide attention and aims to create more realistic humanoid avatars.[13]

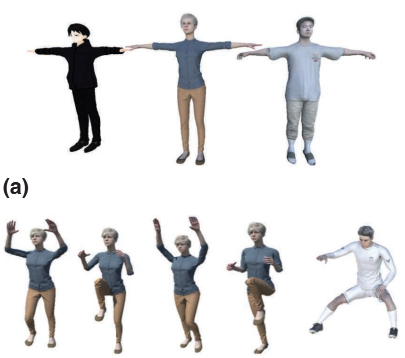

FIGURE 1. The 3D avatar and reconstructed 3D human motion. (a) 3D digital human/avatar. (b) 3D human motion.

3D motion reconstruction

Reconstructing real human motion is an important prerequisite to ensure reliable results of digital human application. The implementation of constructing 3D motion is based on the following techniques.

Inverse kinematic (IK)

Since the humanoid model can be abstracted as the rigid body structure, the body joint parameters can be calculated via the movement of the kinematic chain's end (that is, the end body limb). IK-based design processes provide a basis for 3D motion generation through the manipulation of designers. The process is intuitive and easy to operate. However, the design requires professional knowledge of human movement. The application of IK-based motion generation is still limited and generally applied to small-scale motion simulations, such as the specific movement in a game (for example, holding a bottle, riding bicycles, and so on).

MoCap equipment

Popular solutions to obtain a 3D motion are typically based on MoCap systems. The commonly used tools are the marker-based optical tracking equipment (for example, Vicon[14]), IMU-based MoCap (for example, Xsens[15]), and depth camera (for example, Kinect[16]). The human motion can be accurately built by utilizing professional and mature equipment. However, the costly equipment limited the pervasive development of this technique, and the majority of developed applications were based in the laboratory.

Video

In addition, the rapid development of computer vision and deep learning has recently introduced another type of approach for generating 3D motion. With regard to extracting the skeleton information of a person in a red, green, blue (RGB) video, a deep neural network can be trained to infer the 3D joint position variation and thus generate a 3D motion, for example, the work of VideoPose3D.[13] As the RGB video can be accessed easily by online sources and commercial off-the-shelf devices, the deep model-based method introduced an excellent solution between the IK and MoCap technique. Thus far, some commercial software can be employed to realize a 3D motion transformation from the video, such as DeepMotion.[17] Owing to the convenience and ubiquity of input video, it is believed that the video-based method will dramatically improve the motion generation process with more advanced deep model development.

VIRTUAL SENSORS WITH 3D DIGITAL HUMAN MOTION FOR INTERACTIVE SIMULATION

After obtaining the 3D motion sequence, virtual sensors need to be designed to generate the corresponding synthetic sensor data. In this section, two virtual sensor design processes, namely IMU and distance sensor, are introduced.

Framework of interactive simulation using virtual sensors

Conventional simulation falls into the concept of designing a model of a real system and conducting the experiment with the model.[18] The whole simulation process focused on the high-degree reproduction of the real mechanism. With virtual sensors used, the sensor-based human-activity application's development would become more interactive and intuitive. Specifically, with the 3D motion, the sensor signal generation process is prone to control and understanding. And the user can place the sensor just like in the real world and adjust the human motion easily. And check the result in the virtual environment to iterate the design or directly applied to the real world. The whole framework is presented in Figure 2.

Generating the sensor data

Physics engine systems support the computer software to simulate real-world physics, which normally include rigid body simulation, collision phenomenon, soft dynamics, and so on. Therefore, with the help of the simulated physical environment built by the physics engine, it is possible to design relevant physical sensors for detecting the relevant variables in the virtual environment. The detailed process is shown in Figure 3.

Real sensor systems typically maintain several important parts, including sensor unit, signal processing unit, and digital conversion unit. As the main task is to detect the change of the physical variable, the virtual sensor design focuses on signal-level simulation, which aims to generate related physical variations. Thus, we only focused on the detected variables of the sensor. Figure 3(a) shows a generic implementation pipeline of the virtual sensor design based on Unity3D. Several key steps are introduced as follows.

Key physics simulation

This is the core step of virtual sensor design. The step needs to simulate the detection process in a real sensor, such as the distance value obtained in a virtual environment for a real infrared (IR) distance sensor.

Signal interpolation

Since the reconstructed human motion is frame-by-frame, such a sensing process is also discrete. Therefore, to better utilize the sampled signal from a virtual sensor, it is necessary to interpolate the signal into a continuous function for further analysis and processing.

Resampling

The main application of virtual sensor design is to replace the role of the real sensor in a detection system. Thus, it is also important to map the real sampling frequency to assist a seamless application. After the sensor signal is interpolated, it can discrete the signal again through a flexible sampling as needed.

Filtering

As the main simulation process is based on numerical calculation, noise signals may appear in some situations due to the calculation. Thus, filtering is required in some cases.

IMU

One of the most popular solutions related to the human-motion-based interactive system is the IMU sensor. An IMU provides the basic information from a kinematic aspect. The IMU sensor can obtain the acceleration, angular velocity, and magnetic field. Among them, acceleration and angular velocity information are usually utilized. Therefore, virtual accelerometers and gyroscopes can be designed to generate limb acceleration and angular velocity in 3D avatar motion to simulate the output of a real IMU.

It is easy to obtain the motion coordinates and rotation information of the avatar in virtual space. Therefore, based on the relevant definitions, several kinematic data can be calculated from the displacement and angle [Figure 3(b)]. Specifically, the quadratic

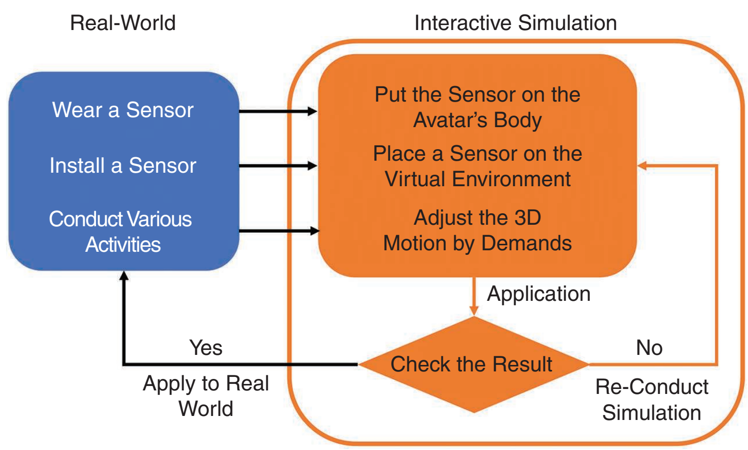

FIGURE 2. The framework for interactive simulation using virtual sensors.

differentiation of displacement is the acceleration, and the primary differentiation of angle is the angular velocity.[19] Note that the displacement and angle information obtained in the virtual space is frame-by-frame, and the discretization of discrete data are usually done by the difference method. Interpolation of discrete data into

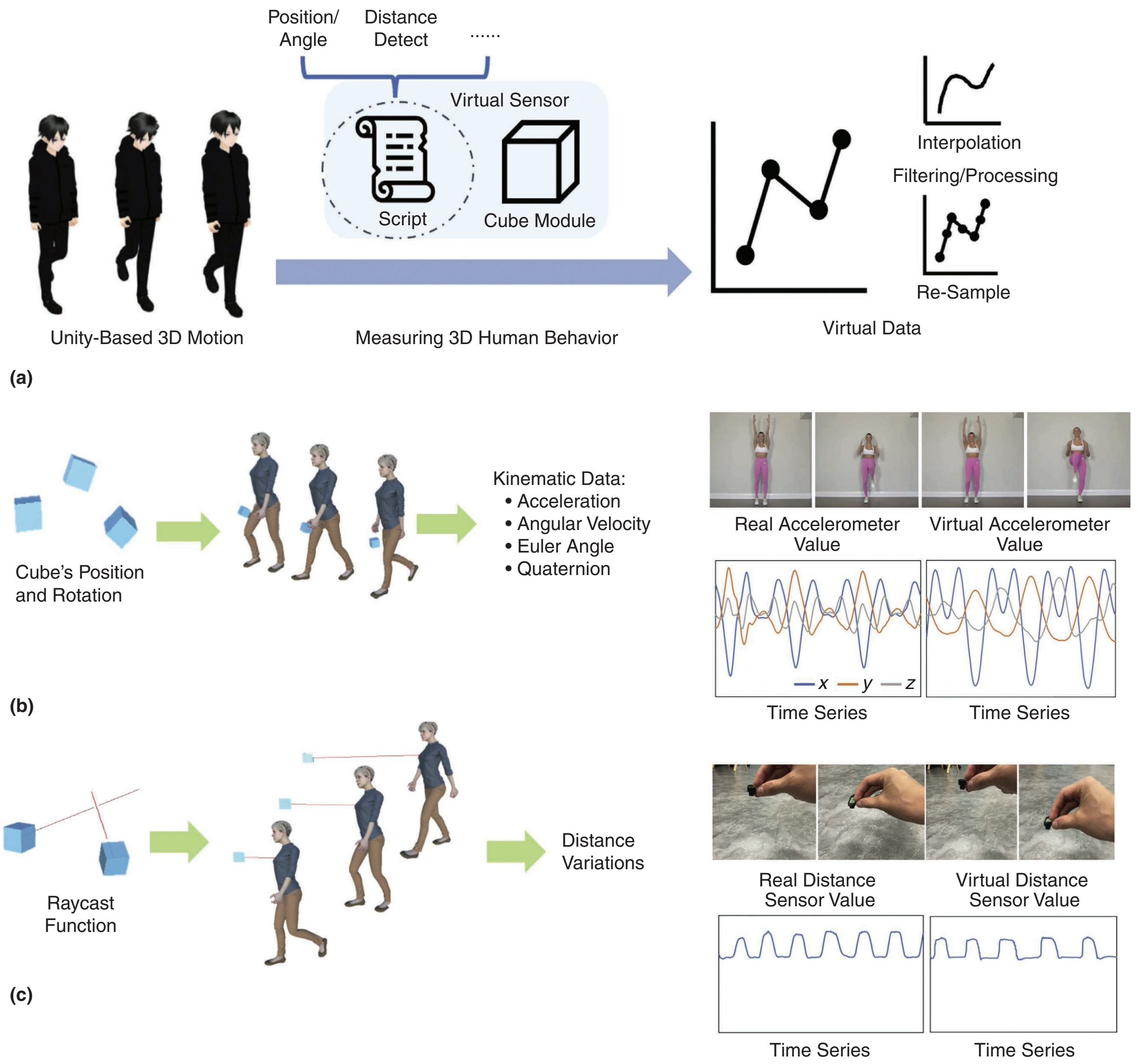

FIGURE 3. The virtual sensor design based on Unity3D. (a) A generic virtual sensor implementation pipeline for measuring 3D digital humans. (b) Virtual IMU sensor for outputting kinematic data via accessing the position and rotation data. An illustration of virtual and real acceleration signals from right upper leg is shown. (c) Virtual distance sensor used to detect the distance variations between the transmitter and the object. An illustration of virtual and real distance signals is shown.

continuous functions can help obtain more accurate differentiation results.

Distance sensor

Using the *Raycast* function in Unity, a real-world IR distance sensor can be simulated to obtain the distance between the target location and the detection point [Figure 3(c)]. When the virtual ray is created, it shoots out in the specified direction, and when it meets colliders, the ray is blocked, and the distance value is returned. This method greatly simulates the detection method of a real IR distance sensor. Although there are various distance sensors in the real world (for example, ultrasonic-based and IR-based), depending on the detection theory, the ultimate goal is to detect distance. Therefore, using a physics engine to simulate an IR distance sensor in virtual space is a feasible solution to obtain the distance between virtual objects.

APPLICATIONS

In this section, two applications of virtual sensors are presented. The first is the development of human activity recognition system associated with ML. The second relates to developing human motion guidance systems, a non-ML application. Both types of applications are real-world applications developed under the concept of virtual sensors and interaction simulation introduced in this article.

Low-dimensional sensor data-based motion recognition system with optimal sensor positions

Human activity recognition (HAR) is a topic that has attracted intense focus as it enables the computer system to understand the human and assist the users' behavior with high efficiency. Generally, the datasets were collected based on the fixed sensors, and the sensor positions have a significant effect on the data distribution. Combining the virtual sensor data with an optimization method can help find the best sensor positions, leading to the highest recognition performance HAR system. Moreover, as the dataset collection would no longer request the real participants to contribute, the low cost is able to allow the more experience-less designer to be engaged and attempted. Thus, the whole development process can be more flexible and interactive. The user could access any interesting sensor position to check the recognition performance. The following cases show the design of HAR systems using 3D digital human motions and the previously introduced sensor data collection process via virtual sensor design.[20] In these cases, the data for training the HAR classifier are collected from 3D digital human motion in the virtual space. After finding the best positions, the classifiers are trained to recognize real-world activities by data augmentation.[21]

Figure 4(a) shows the wearable IMU sensor attached to the 3D digital human. The input motions were captured from the IMU MoCap suite, namely, Xsens MVN. Three people with five types of activities, including *walking*, *running*, *standing*, *going upstairs*, and *going downstairs*. A total 90 s of virtual acceleration data were collected and segmented by a 2-s window. After three times of augmentation, the data were segmented to extract the time- and frequency-domain features and then used to train a support vector machine (SVM) classifier. For real testing data, the same length acceleration data were utilized, and 89.85% accuracy could be obtained while using the classifier trained by virtual data to recognize the real activity data (based on three sensor positions *chest*, *right shoulder*, *head* after optimization).

Figure 4(b) shows a virtual distance sensor being used as a wearable to detect the distance variations, which are caused by the user's motion, between the on-body distance sensor and the ground. The distance sensor can be attached to the lower limbs of the body with a different transmission angle. Three people were invited to perform the exercise activities for 60 s, including *heel up/down*, *squat*, and *hip stretch*. Xsens MVN was used to reconstruct the 3D motion in the virtual environment. The data processing, feature extraction, and classifier training followed the aforementioned method. In the real world, real distance sensors (GP2Y0A21YK0F, Sharp) with a 60-Hz sampling rate were selected, and 91.25% accuracy could

GENERALLY, THE DATASETS WERE COLLECTED BASED ON THE FIXED SENSORS, AND THE SENSOR POSITIONS HAVE A SIGNIFICANT EFFECT ON THE DATA DISTRIBUTION.

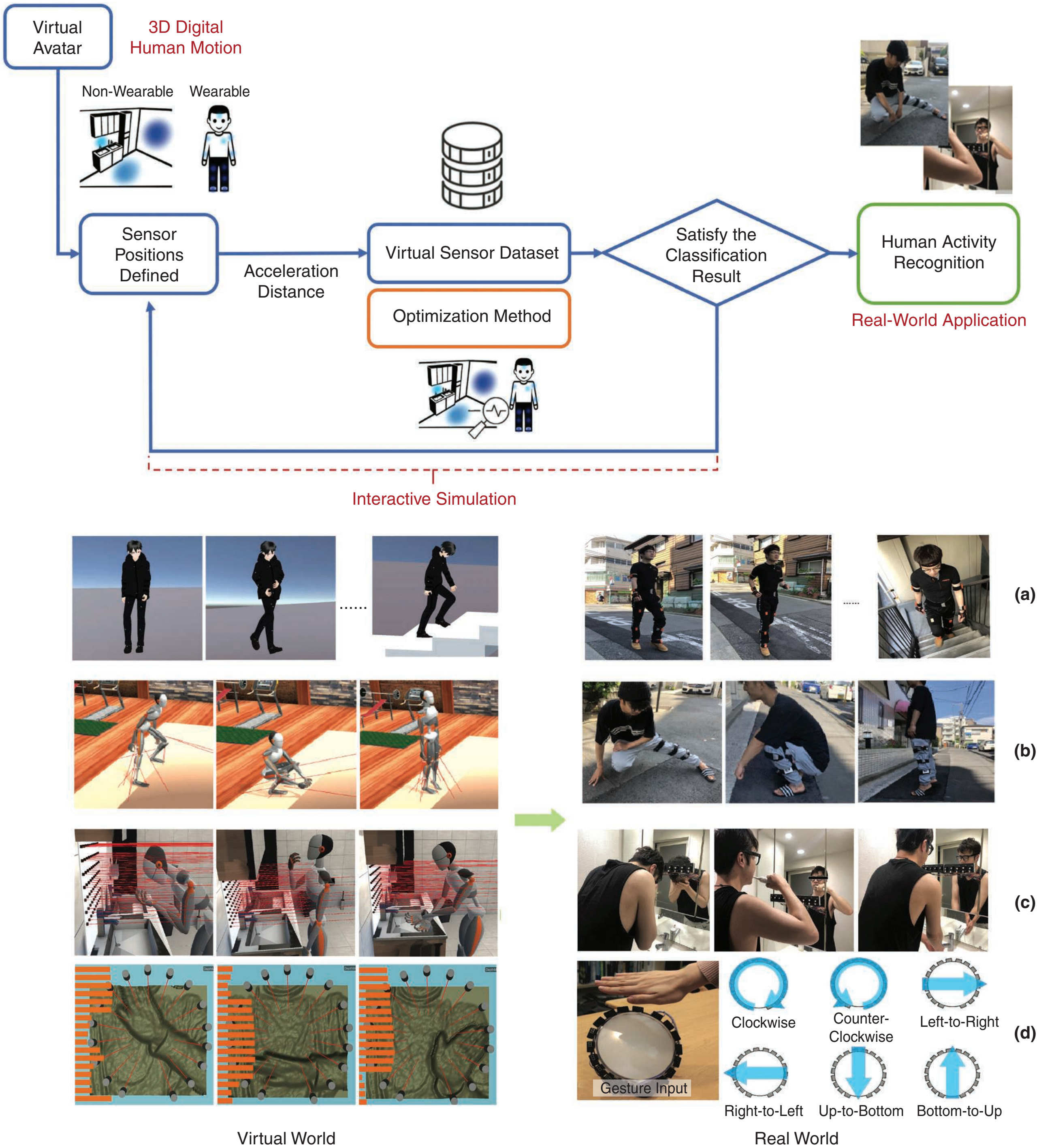

FIGURE 4. Virtual sensor detection from 3D human motion applied to HAR system design. (a) Using virtual wearable IMUs to design a daily activity recognition system. (b) Using wearable virtual distance sensors to recognize the exercises. (c) Ambient-based virtual distance sensors used to identify the activities in the bathroom. (d) Support software to assist the hand gesture recognition interface with distance sensors.

be achieved while using the virtual data-related classifier to recognize the real activity (based on two sensor positions *left lower leg-left* and *left upper leg-back* after optimization).

Conventional dense sensing normally adopts large amounts of sensors and datasets to improve the robustness of the system. The required area is generally a whole ceiling or a large part of the wall. Using the virtual distance sensor can more easily provide the guidance of the sensor's position and number to recognize activity. We assumed the applied scene was the bathroom of the home, and the system aimed to track the daily activities that occur in the bathroom, including the *washing hands*, *washing face*, and *brushing teeth*. Therefore, in the virtual environment, 70 virtual distance sensors were placed on the wall above the washbasin, as shown in Figure 4(c). The 70-sensor area was then divided into several subsensor boards (24 subboards, cf. the blue and red blocks). Three participants were recruited, and a total 90 s of the related activities were reconstructed by Xsens MoCap. The sampling rate was set as 25 Hz and segmentation length was 2 s. The collected virtual distance sensor data were aligned and converted into grayscale figures. The extracted texture features via Gabor filter were then imported into the SVM classifier for training.

In the real world, the IR distance sensor (GP2Y0A21YK0F, Sharp) and the Arduino chip (ARDUINO PRO MINI) were utilized. After the real and virtual coordinates were mapped, the testers were requested to stand in front of the real sensor boards and conduct the corresponding activities for 60 s. Following the same aforementioned approach, the data were converted into grayscale figures. The accuracy of using a virtual-data-trained classifier to recognize the real activity could reach 90.69% (based on ten distance sensors used after optimization).

In addition to the recognition and interaction of full-body motion, hand motion is also a topic that has garnered increased focus in the human–computer interaction field. With interface systems evolving toward being natural and efficient, recognizing gestures has played an important role in this development. The traditional prototyping process is affected by the position of the utilized sensor and follows the approach of empirical trial and error, which requires the expertise of the developer. Using the virtual distance sensor can assist such a prototyping process for hand-gesture recognition interface development. In the virtual sensor-based approach, software can be developed to support the virtual distance collection caused by virtual hand motion input.[22] It also fuses the optimization process and helps reduce the discipline barrier for development. The output of the designed software is the specific sensor placement and classifier. In the real world, the real distance sensors imitate the corresponding sensor position and recognize the human's hand motion through the generated ML model [as shown in Figure 4(d)].

The input hand motion was recorded by Microsoft Kinect. Subsequently, the captured 3D data were reconstructed by the mesh in the virtual environment. The mesh structure was built in a high dimension, and the related coordinate system was transferred from the coordinate system of Kinect. After the hand motion was imported, the virtual distance sensors could then be placed by considering the practical prototype situation. Since various distance sensors can be deployed, the detected distance value is converted into a grayscale image for recognition. After the normalization, the matrix is converted into a grayscale image, which is then input into a CNN classifier for training.

To test the performance, a real IR distance sensor (GP2Y0E02A, Sharp) was employed in the real world. Seven people were recruited to test the system. Each individual's data were used to generate the classifiers, and the remaining six people's data were employed to test the performance. To adapt the virtual-data-trained classifier into the real-world domain, transfer learning was utilized with a small amount of real data to fine-tune the fully connected layer of pretrained CNN models by virtual data. The recognition accuracy could reach more than 90% with the transfer learning used.

Customized motion-learning system combined the digital human and auditory feedback

Motion-learning system has contributed to human daily exercise and health to a large extent. Conventional motion-learning systems typically employ video tutorials to assist the user in studying the required motion for exercise training, rehabilitation, and other activities. However, the limited information restricts the learning effect, and various users may cause different study results for a given motion. In addition, most systems are utilized based on a predefined motion, exhibiting low flexibility and difficulty adjusting to different usage conditions. Thus, finding a new solution based on an efficient, convenient system and improving the flexibility of a learning system can benefit more motion learning applications and engage more users.

Drawing from the concept of the virtual sensor, a novel motion-learning system, *VoLearn*, was developed based on a 3D avatar (Figure 5). *VoLearn* fuses several key components to contribute to a customized motion design and provide effective feedback during the learning process.[23] Specifically, it converts the pervasive 2D motion video into 3D motion with a state-of-the-art deep learning model. For a given motion 3D file, an interface is developed to allow the user to fine-tune and build a secondary design. The speed and amplitude of the motion can be adjusted according to the user, thereby allowing the different motion combinations to form a new motion. After the motion is designed, the system conducts an online analysis of the virtual motion data to recommend a sensor position for real-world study. In the real world, the user is required to wear the personal smartphone on the designated body parts, and the smartphone-based application will analyze the user's behavior to produce auditory feedback for the user. The interactive feedback includes the error information regarding the user's amplitude and speed, which can help the user learn the motion more accurately. *VoLearn* can help the user learn a motion more accurately and

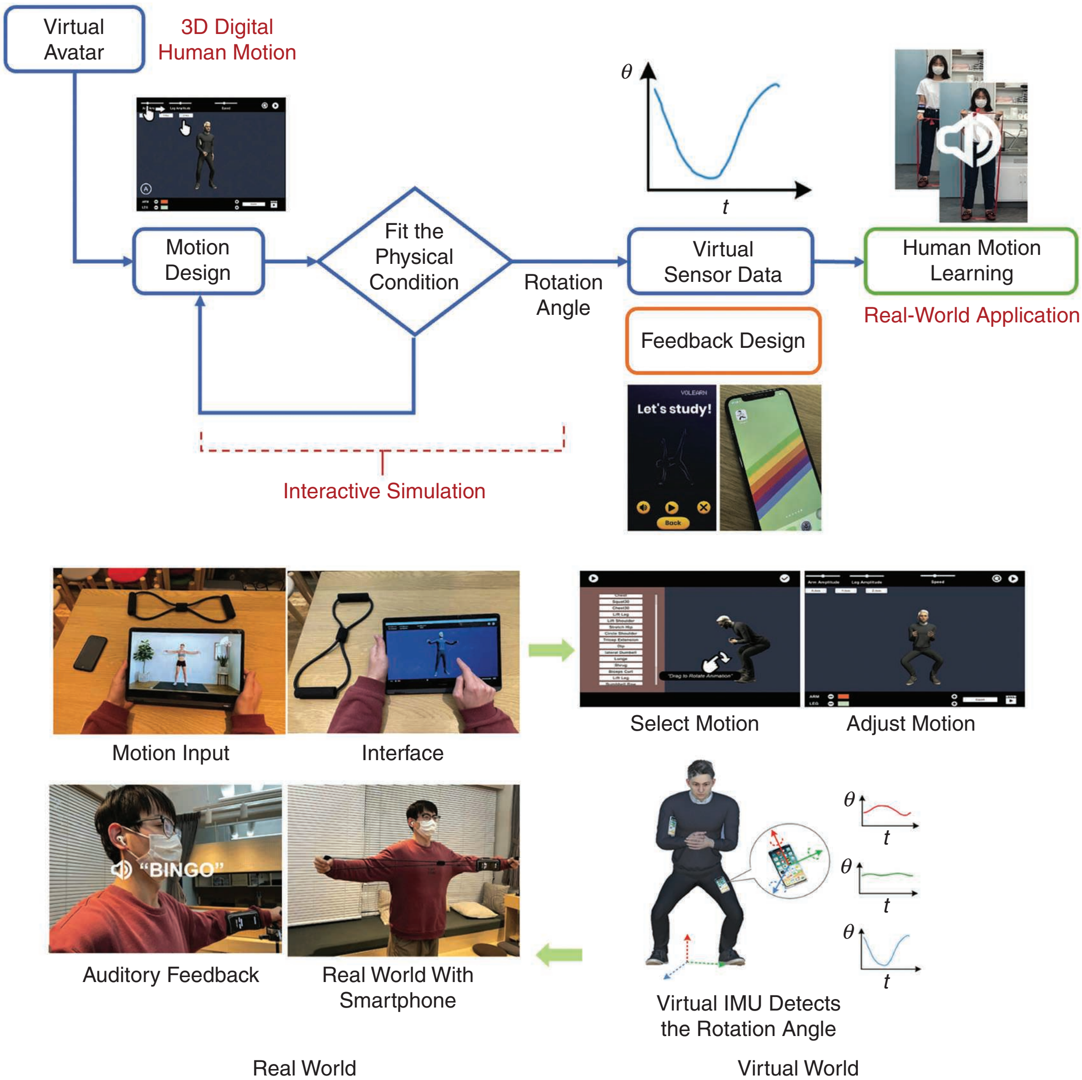

FIGURE 5. Virtual IMU-based detection on 3D human, applied to customized motion-learning system.

proves the effectiveness of the corresponding sports training, rehabilitation, and therapy scenarios.

FUTURE AND CHALLENGES

Collecting sensor data from 3D digital human motion by virtual sensors provides a novel development paradigm for traditional real sensor data-based systems, especially low-dimensional sensor data ML development. Unlike the traditional real dataset collection as a premise, the presented virtual sensor-based detection greatly decreases development costs and alters the characteristics of conventional human activity-related applications.

Accurate and accessible 3D motions

Thus far, to obtain accurate 3D motion to create sensor data, high-performance MoCap equipment could be employed. Nevertheless, the laborious and costly process has limited pervasive applications. Other methods, such as using the deep learning model to convert a 2D video to a 3D motion, have become mainstream. The low-cost and wide video source makes 3D motion files increasingly accessible. However, converted 3D motion still provides limited performance due to the learning-model-based mechanism. For example, if the person in the video is obscured or the video quality is low, the converted 3D person's movement is poor. Moreover, the converted 3D motion is better for motion videos with a large range of motion. Therefore, the next key focus is on obtaining high-performance 3D character motion sequences cost-effectively. With the development of deep-learning vision models, it is believed that a wide range of high-performance 3D human motion sequences will be highly accessible. Using virtual sensor-based detection on these 3D motions can lead to novel real-world applications related to human motion.

Virtual sensor design hints

Two virtual sensor designs, namely, the IMU and the distance sensor, are presented in this article. As introduced before, the realization of the virtual sensor signal mainly depends on the physics engine. With further development of the physics engine, more types of sensors can be simulated and developed, such as light, radar, and flex sensors. The use of multiple types of virtual sensors related to virtual-reality sensor fusion algorithms and information intelligence applications will be expected. Figure 6 shows potential virtual sensor design hints. Since the design of a virtual sensor based on a physical engine does not simulate the sensor at the principle level, the design is application-oriented in its thinking. Defining the variables detected by the sensor is the focus of designing such virtual sensors. For example, the IMU sensor can detect angular velocity, but it can also be used to measure the degree of joint flexion of a limb during motion. Therefore, for the former virtual sensor design, angular velocity calculation is required, while for the latter design, Euler angles can be directly accessed for simulation. Table 1 gives several potential virtual sensor developments.

More realistic virtual sensor signal

A major issue limiting the application of virtual sensors is the realism of the signals. Since the detection is based on reconstructed 3D human motion sequences, ensuring that the signals generated from virtual sensors are as close as possible to those generated under real human motion remains a major difficulty. There are many factors that cause this difficulty, including the 3D avatar size, the motion characteristics of different real characters, the interference of other physical

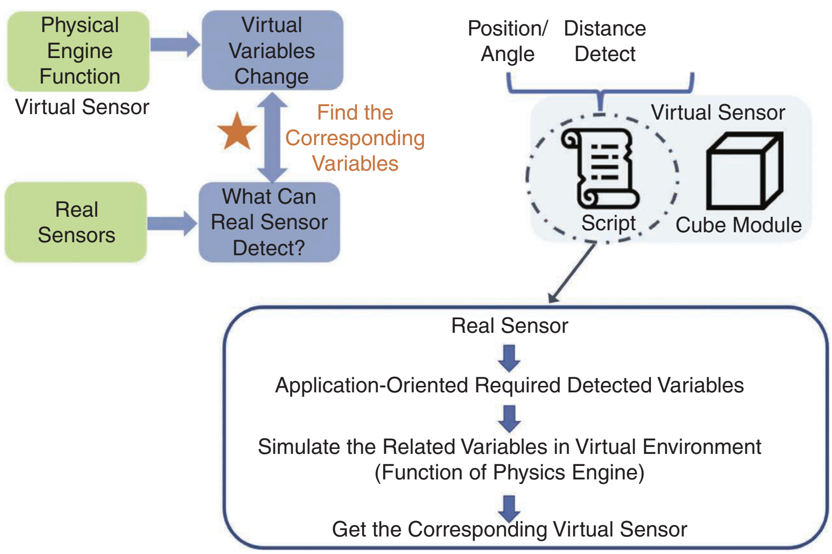

FIGURE 6. The virtual sensor design hints.

TABLE 1. Potential virtual sensors design.

Function in the game engine	Variables in the virtual environment	Variables in the real world	Sensors in the real world
Light	Brightness value	Brightness value	Light sensor
Soft-body simulation	Shape changed	Shape changed	Flex sensor
Collider	If there was a collision	If two objects touch	Capacitive sensor
Object's rotation	Bending degree	Bending degree	Flex sensor IMU sensor
...	...	...	...
Raycast	The distance between the transmitter and the obstacle	The distance value	Distance sensor
Object's position and rotation	The acceleration, angular velocity, Euler angle, quaternion	Acceleration, angle, quaternion, and so on	IMU sensor

characteristics in the real world (magnetic field, multipath effect, and so on), and the soft tissue of the wearing situation. So far, to minimize the difference between virtual data and real data, the number of activity recognition categories is usually limited, for example, three to four, to ensure that the classifier trained on virtual data can recognize real activity data with better performance. To solve such problems, it is necessary to continuously improve the fidelity of the overall environmental factors in the virtual world. On the other hand, it is also possible to use a small amount of the real world to retrain the classifier from a data perspective, such as transfer learning, so that it can be better applied in real environments.

However, the latter approach still requires a certain amount of real data and brings little improvement compared to the traditional development paradigm. Therefore, the main research results at this stage are still focused on low-dimensional sensor virtualization and minimizing the difference between virtual and real data distribution. To advance promising developments, exploring the virtual sensor data-specified ML development method, such as virtual-to-real domain-invariant features, can also be focused on. With the development of related research, virtual environment applications will become increasingly complete. While traditional techniques have mainly concentrated on reproducing real spaces in virtual spaces, the methods presented in this article will complement another idea of using data collected from virtual spaces to improve real-world sensor system applications. In addition to the immersion brought by virtual reality devices at this stage, systems developed using virtual sensors will obtain another sense of seamlessness in the virtual–real world and create an entire cyberphysical world. Moreover, without the heavy requirement of real datasets in traditional ML systems, ML systems developed using virtual sensors will be lightweight, flexible, and will create more personalized services for users.

Traditional sensor-based human activity-related applications are limited by the costly real data required, making development time-consuming and laborious. This article presented a novel sensor data collection method for human activity. With the help of a physics engine, a virtual sensor can be designed to measure 3D human motion and thus generate corresponding sensor data. With this approach, the traditional real-world measurement of human-related activity using sensors can be transferred to the virtual environment, significantly reducing development costs and enabling interactive simulation. This approach can bring more flexibility and customization to HAR systems.

ACKNOWLEDGMENT

This work was supported by JST PRESTO Grant JPMJPR2134.

REFERENCES

1. Y. Celik, S. Stuart, W. L. Woo, E. Sejdic, and A. Godfrey, "Multi-modal gait: A wearable, algorithm and data fusion approach for clinical and free-living assessment," *Inf. Fusion*, vol. 78, pp. 57–70, Feb. 2022, doi: 10.1016/j.inffus.2021.09.016.
2. C. S. Qiu et al., "Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges," *Inf. Fusion*, vol. 80, pp. 241–265, Apr. 2022, doi: 10.1016/j.inffus.2021.11.006.
3. T. T. Pham and Y. S. Suh, "Conditional generative adversarial network-based regression approach for walking distance estimation using waist-mounted inertial sensors," *Inf. Fusion*, vol. 71, May 2022, Art. no. 2510113, doi: 10.1109/TIM.2022.3177730.
4. M. Alzantot, S. Chakraborty, and M. Srivastava, "SenseGen: A deep learning architecture for synthetic sensor data generation," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, 2017, pp. 188–193, doi: 10.1109/PERCOMW.2017.7917555.
5. J. Wang, Y. Chen, Y. Gu, Y. Xiao, and H. Pan, "SensoryGANs: An effective generative adversarial framework for sensor-based human activity recognition," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489106.
6. X. A. Li, J. Luo, and R. Younes, "ActivityGAN: Generative adversarial networks for data augmentation in sensor-based human activity recognition," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, 2020, pp. 249–254, doi: 10.1145/3410530.3414367.

ABOUT THE AUTHORS

CHENGSHUO XIA is a lecturer at the Advanced Manufacturing Technology Innovation Center, Guangzhou Institute of Technology, Xidian University, Guangzhou 510555, China. His research interests include ubiquitous computing and human–computer interaction. Xia received a Ph.D. from Keio University. Contact him at csxia@keio.jp.

YUTA SUGIURA is an associate professor in the Department of Information and Computer Science, Keio University, Yokohama 223-8522, Japan. His research interests include user interfaces, ubiquitous computing, and medical–engineering collaboration. Sugiura received a Ph.D. from the Graduate School of Media Design at Keio University. Contact him at sugiura@keio.jp.

7. H. Kwon et al., "IMUTube: Automatic extraction of virtual on-body accelerometry from video for human activity recognition," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–29, Sep. 2020, doi: 10.1145/3411841.
8. K. Ahuja, Y. Jiang, M. Goel, and C. Harrison, "Vid2Doppler: Synthesizing Doppler radar data from videos for training privacy-preserving activity recognition," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–10, doi: 10.1145/3411764.3445138.
9. M. Mochimaru, "Digital human models for human-centered design," *J. Robot. Mechatronics*, vol. 29, no. 5, pp. 783–789, Oct. 2017, doi: 10.20965/jrm.2017.p0783.
10. "Bring your vision to life." Unity 3D. Accessed: Jan. 2023. [Online]. Available: https://unity.com/
11. "Get animated." Mixamo. Accessed: Jan. 2023. [Online]. Available: https://www.mixamo.com/
12. "Open source tool for making 3D characters." MakeHuman Community. Accessed: Jan. 2023. [Online]. Available: http://www.makehumancommunity.org/
13. D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 7745–7754, doi: 10.1109/CVPR.2019.00794.
14. "Motion capture systems." Vicon. Accessed: Jan. 2023. [Online]. Available: https://www.vicon.com/
15. "Motion tracking system." Xsens. Accessed: Jan. 2023. [Online]. Available: https://www.xsens.com/
16. *Kinect for Windows*. (Aug. 2022). Microsoft. [Online]. Available: https://docs.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows
17. "AI motion capture & body tracking." DeepMotion. Accessed: Jan. 2023. [Online]. Available: https://www.deepmotion.com/
18. J. Banks, "Introduction to simulation," in *Proc. 31st Conf. Winter Simul., Simul. Bridge Future*, 1999, pp. 7–13, doi: 10.1145/3410530.3414367.

19. A. D. Young, M. J. Ling, and D. K. Arvind, "IMUSim: A simulation environment for inertial sensing algorithm design and evaluation," in *Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sens. Netw.*, 2011, pp. 199–210.
20. C. Xia and Y. Sugiura, "Optimizing sensor position with virtual sensors in human activity recognition system design," *Sensors*, vol. 21, no. 20, pp. 6893–6903, Oct. 2021, doi: 10.3390/s21206893.
21. T. T. Um et al., "Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks," in *Proc. 19th ACM Int. Conf. Multimodal Interact.*, 2017, pp. 216–220, doi: 10.1145/3136755.3136817.
22. C. Xia, A. Saito, and Y. Sugiura, "Using the virtual data-driven measurement to support the prototyping of hand gesture recognition interface with distance sensor," *Sensors Actuators A, Phys.*, vol. 338, no. 2, Feb. 2022, Art. no. 113463, doi: 10.1016/j.sna.2022.113463.
23. C. Xia, X. Fang, R. Arakawa, and Y. Sugiura, "VoLearn: A cross-modal operable motion-learning system combined with virtual avatar and auditory feedback," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 2, pp. 1–26, Jul. 2022, doi: 10.1145/3534576.

Digital Object Identifier 10.1109/MC.2023.3326620