Sensors and Materials, Vol. 36, No. 1 (2024) 291–303
MYU Tokyo

S & M 3519

Real-time Fall Prediction Using Kinect Sensor with Random Sample Consensus Algorithm

Po-Tong Wang,[1] Yu-Jen Chen,[2] and Jia-Shing Sheu[2*]

[1]Department of Electrical Engineering, Lunghwa University of Science and Technology,
No. 300, Sec. 1, Wanshou Rd., Guishan District, Taoyuan 333326, Taiwan
[2]Department of Computer Science, National Taipei University of Education,
No. 134, Sec. 2, He-Ping East Road, Da-an District, Taipei 106, Taiwan

(Received July 11, 2023; accepted December 18, 2023)

Keywords: Kinect, fall detection, random forest classifier, random sample consensus (RANSAC), Hough transform

Falls are a leading cause of unintentional injury and death and pose a significant risk to individuals living alone, particularly older adults. The World Health Organization reports that approximately 684000 fatal falls occur annually, and that timely fall detection is essential to enable prompt treatment and mitigate harm. To address this challenge, we propose a novel in-home fall detection system that uses the Kinect V2 sensor instead of a wearable device. The system can promptly alert a person's caregivers or family members when it detects a fall, enabling timely assistance. The Kinect V2 sensor captures depth frames and performs skeleton tracking, and body parts in single-depth image pixels are identified using a random forest classifier. v-disparity images, the Hough transform, and the random sample consensus algorithm are used to identify the floor in depth images, and falls are predicted by analyzing the vertical accelerations of 10 joints in the human skeleton. The system can issue an alarm up to 0.5 s before a fall, enabling preventive measures to be taken. In experiments, the proposed system was effective in accurately predicting falls and hence, has potential to improve fall risk management in caregiving environments.

1. Introduction

Falls represent a significant hazard, particularly for older adults living alone. A fall may be life-threatening; hence, rapid detection and immediate assistance are crucial. Present fall prediction or detection methods primarily use threshold-based or machine-learning approaches. In this research, we introduce a novel vision system and algorithm to predict falls from human joint acceleration and the relative distance between the joints and the floor. This innovation is aimed at expediting assistance by predicting and alerting appropriate individuals to falls as they occur.

Our proposed system employs the Kinect V2 sensor to capture depth images. These images undergo a series of transformations, including image coordinate conversion into v-disparity

*Corresponding author: e-mail: jiashing@tea.ntue.edu.tw
https://doi.org/10.18494/SAM4648

ISSN 0914-4935

images, binarization, Hough transformation, conversion into v-disparity images, and application of the random sample consensus (RANSAC) algorithm. In this process, floor parameters are extracted and used with Kinect V2's body recognition to determine joint acceleration and predict falls.

The aim of this study is to develop a reliable fall prediction system that employs a low-cost depth camera. The system should significantly improve healthcare outcomes and reduce costs. Falls by older adults are increasingly prevalent, and the healthcare costs associated with fall-related injuries are increasing. Hence, the economic burden of falls is substantial, and cost-effective solutions such as our proposed system could significantly mitigate these expenses.

The remainder of this study is organized as follows. First, in Sect. 2, the pertinent literature is reviewed, followed by a detailed explanation of our methodology and the system architecture in Sect. 3. In Sect. 4, the experiment is outlined and we present our findings. We discuss the conclusions derived from our research in the final section.

2. Literature Review

In this section, we provide a comprehensive review of existing literature on fall detection systems. Various techniques and technologies have been leveraged to develop these systems, broadly classified as vision-based, non-vision-based, and hybrid systems incorporating a neural network.

Non-vision-based fall detection systems typically use wearable devices to capture human movement data. These devices often incorporate gyroscopes or accelerometers to identify human posture. For instance, in one study,[1] falls were differentiated from other human movements by analyzing the body's vertical and horizontal velocities and using a threshold algorithm to detect falls. In another study,[2] a gyroscope was employed to record angular momentum and developed a threshold algorithm to detect a fall. Additionally, accelerometers have been employed to measure human acceleration for fall detection.[3] Integrated gyroscope and accelerometer data have also been integrated to recognize postures and identify falls[4] more accurately. These non-vision-based systems reveal that diverse methods can be employed for fall detection and highlight the potential for effective fall detection through an innovative combination of technology with data analysis.

Wearable devices can be inconvenient for users; hence, vision-based fall detection systems, which are less intrusive than systems using wearable devices, have also been developed. Before depth cameras became widely available and affordable, several groups analyzed RGB camera images to detect falls.[5–7] However, using a single camera has numerous limitations, such as blind spots and a restricted angle; multicamera systems were thus proposed to improve fall detection.[8,9] Microsoft's introduction of Kinect marked a significant shift in the field. Kinect contains an affordable RGB-depth (RGB-D) camera, which has been applied in Kinect-based fall detection systems.[10–12] One key benefit of depth image analysis is that a user's identity and privacy can be protected because the captured images do not reveal facial details but provide rich data for scene and body analysis. Furthermore, unlike conventional cameras, the lighting

conditions do not affect the depth of camera images. Studies on methods for human body segmentation within a scene to identify various features, such as velocity, height, and weight, which could be applied for fall detection, have been reported.[10,11] Obstacles in a scene may reduce detection performance. To overcome this challenge, Mundher and Zhong proposed equipping a mobile robot with the Kinect system[12] to increase the system's flexibility and range in complex environments. To improve the accuracy and performance of fall detection methods, a hybrid approach combining wearable devices with Kinect has been presented in an embedded platform.[13–15] Kinect and the wearable device captured depth and motion data, respectively, and the machine-learning method, support vector machine (SVM), was used for data analysis and fall detection.

Machine learning has enabled further improvement of the predictive capabilities and performance of fall detection systems. The fall detection performance of deep-learning techniques, such as the recurrent neural network, and machine-learning methods, such as SVM and naive Bayes, were compared for data captured by wearable devices.[16] The results highlighted the potential of machine learning for improving fall detection systems. A model combining three-dimensional (3D) convolutional neural networks (CNNs) with long short-term memory was proposed for the detection of falls in kinematic video data; these advanced machine-learning techniques could effectively process complex types of data to predict and detect falls accurately.[17] Predicting falls before they occur could significantly increase the timeliness of emergency response. One fall prediction method[18] comprises a combination of a CNN with class activation mapping (CAM). The system could highlight the class-specific features relevant to fall prediction as a heat map; CAM thus revealed the mechanism of fall prediction. The study showed that machine learning could detect and predict falls, and such a forecast can increase the timeliness and effectiveness of subsequent medical intervention.

3. System Architecture and Methodology

In this study, we use an algorithmic approach to develop a home fall detection system. The system includes depth frames and skeleton tracking features. The system can predict falls and provide a timely warning by analyzing factors such as acceleration and the distance between a person and the floor. Figure 1 depicts the system's architecture, generated using integrated computer-aided manufacturing definition for function modeling, called IDEF0. IDEF0 is a modeling language commonly employed in software engineering[19–21] and is based on the functional modeling language, structured analysis, and design technique.[22] Three key techniques were used to predict falls: v-disparity-based floor detection shown in Fig. 2, body part recognition, and fall prediction based on acceleration and distance. The floor detection module uses v-disparity images to analyze depths and identify the floor region. This information is vital for accurately assessing a person's position relative to the floor. The body part recognition component recognizes and tracks specific body parts using the skeleton-tracking features provided by the Kinect V2 sensor. Finally, the fall prediction module uses the collected data, including acceleration and distance measurements, to make informed predictions regarding fall events.

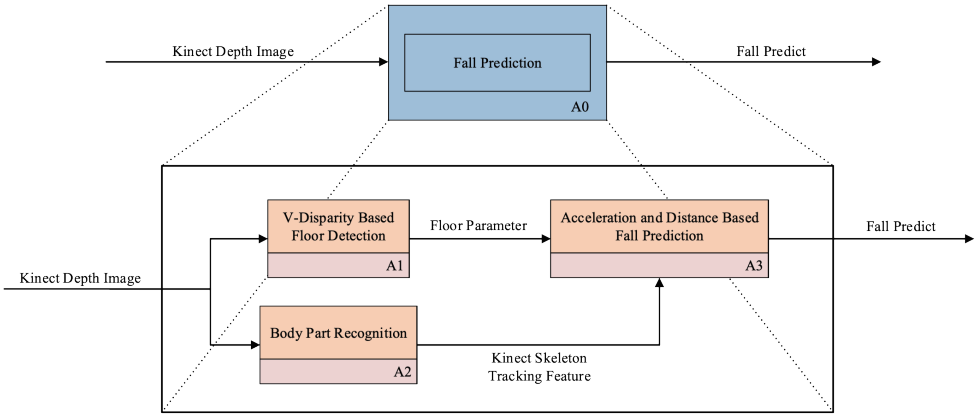

Fig. 1. (Color online) Architecture of the proposed system.

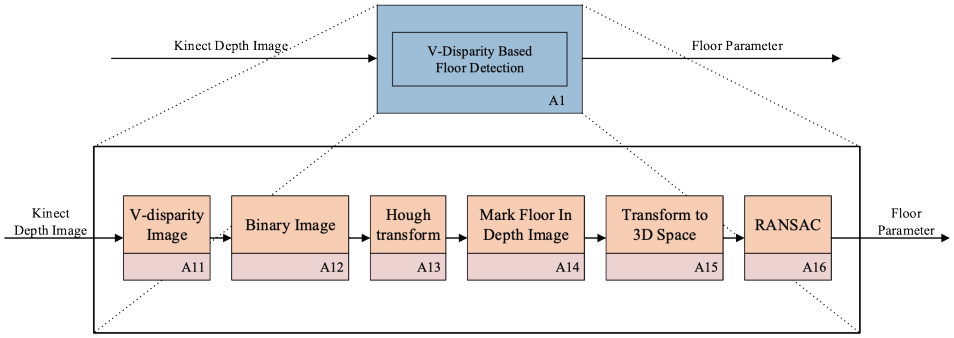

Fig. 2. (Color online) v-disparity-based floor detection.

These components are combined to create an effective and reliable fall detection system. By leveraging advanced sensing technologies and algorithmic analysis, the system can improve safety and reduce the risk of falls in various settings.

Stereo images can be used for reliable obstacle detection.[23] We use depth images captured by the Kinect V2 sensor—which comprises an infrared (IR) camera and IR projector, forming a stereo pair—for floor detection. Disparity d, a measure of the differences between pixels in stereo images, is calculated using Eq. (1), where b is the horizontal distance between the cameras in meters, f is the focal length in pixels, and z is depth in meters. The values of b and f are set to 5 and 367.0094, respectively. A v-disparity image is composed of straight lines and oblique lines. The conversion method is to calculate the number of same disparities on the axis of the original disparity image and accumulate them, as shown in Fig. 3. For a given row in the original disparity image, the count of a disparity value in that row is listed in the corresponding row and column of the v-disparity image. For example, the disparity value 1 occurs twice in the first row of the original image in Fig. 3; hence, the second column (disparity value 1) of the first row of the v-disparity image contains the value 2. Through this process, the disparity image is converted

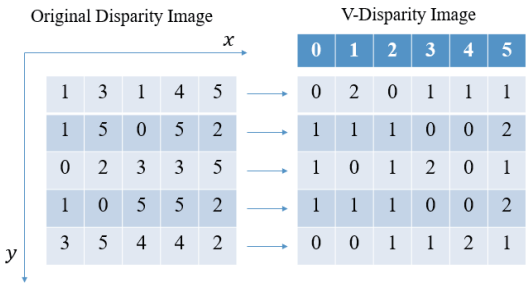

Fig. 3. (Color online) Conversion from disparity image to v-disparity image.

into a v-disparity image. The v-disparity image represents the differences in the disparity between rows and therefore offers valuable insight into depth variation and can be used to detect the floor region. Calculating the v-disparity image from the depth information provided by the Kinect V2 sensor is critical for the subsequent steps of the fall detection method for real-time fall prediction.

$$d = \frac{b \times f}{z} \tag{1}$$

Thresholding is an image segmentation method in which grayscale images are converted into binary images. Applying a threshold to the v-disparity image enables the separation of relevant features of interest from the background. Subsequent operations, such as the Hough transform, can achieve more accurate and robust detection of relevant features, such as lines or curves, on the simplified, binarized v-disparity image, ensuring that the identification and segmentation of the floor region are accurate. The Hough transform is an image-processing technique widely used for feature extraction and typically detects straight image lines. The algorithm uses a point-slope formula to convert the image pixels from Cartesian into polar coordinates. A voting mechanism determines the parameters ρ and θ representing the detected lines. In Fig. 4, ρ represents the shortest distance between the origin and the line, and θ is the angle between this line and the x-axis. Each line in the image can be represented in the form shown in Eq. (2). Although an infinite number of such possible (ρ, θ) pairs passing through a given point can be defined in Fig. 5, in practice, they are restricted to $\rho > 0$ and $0 \leq \theta < 2\pi$ in Eq. (3). When two points intersect in ρ–θ or polar coordinates, as shown in Fig. 6, they belong to the same line. We can accurately identify and display lines in the image by converting these points back to Cartesian coordinates. In sum, the Hough transform detects lines by identifying points of intersection in the polar coordinate space. The resulting line detection and extraction are robust and used in the subsequent floor detection and fall prediction stages.

$$\rho = x\cos\theta + y\sin\theta \tag{2}$$

$$\rho > 0 \text{ and } 0 \leq \theta < 2\pi \tag{3}$$

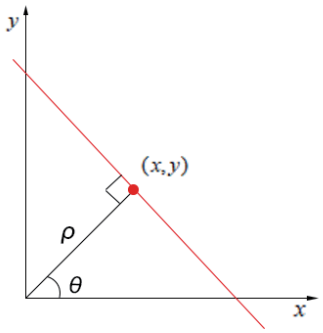

Fig. 4. (Color online) Hough transform.

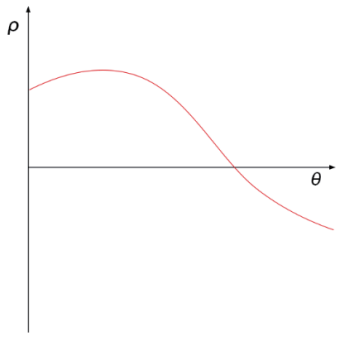

Fig. 5. (Color online) Corresponding line in the polar coordinate plane.

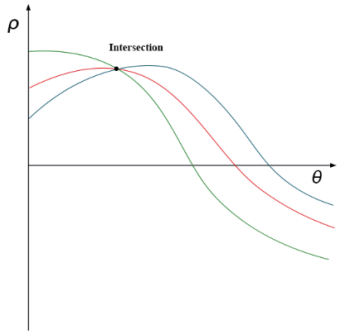

Fig. 6. (Color online) Several coordinate curves in the polar plane.

Identifying in-depth floor images begins with converting the v-disparity image into a depth image. In this process, the v-disparity image, derived from the original depth image, is reversed to reacquire the depth image. This allows the floor pixels to be marked by projecting the segmented slanted lines, extracted through the Hough transform, back onto the depth image. This step transforms the identified two-dimensional floor pixels into 3D camera space. This transformation is facilitated by the Kinect for Windows Software Development Kit 2.0 method, which seamlessly converts 2D pixel coordinates into corresponding 3D points. The result is an accurate spatial representation of the floor within the 3D camera space, thus enabling the precise delineation of the floor region in the depth image.

RANSAC is a robust regression algorithm for estimating the parameters of a model from a dataset that may contain outliers.[24] The RANSAC algorithm randomly samples subsets of size n from the dataset, where n is the minimum number of samples required to estimate the model parameters. It then distinguishes *inliers* from *outliers* for all points in the dataset, and the numbers of inliers for the current model and the best previous model are compared. The model parameters and the maximum number of inliers achieved are recorded. This process is continued until a termination criterion is satisfied.

The termination criterion for RANSAC is typically a minimum number of required samples. Let the proportion of inliers in the dataset be denoted as t in Eq. (3); the probability of selecting

at least one outlier among the sampled points is $1\text{-}t^n$. If the probability of convergence after iteration k is p in Eq. (4), the maximum number of iterations k can be expressed as Eq. (5).

The RANSAC algorithm is used in the floor identification module to compute the plane equation (6) by sampling three 3D floor pixels. The algorithm calculates the distance from each floor point to the proposed plane and determines a suitable distance threshold. Points at distances below the threshold are considered inliers. The process is iterated to select the optimal plane representing the floor. By employing the RANSAC algorithm, our method robustly estimates the floor parameters, even in the presence of outliers. This ensures accurate detection and tracking of the floor, which is crucial for effective fall detection and prevention.

$$t = \frac{n_{inliers}}{n_{inliers} + n_{outliers}} \tag{4}$$

$$p = 1 - \left(1 - t^n\right)^k \tag{5}$$

$$k = \frac{\log\left(1-p\right)}{\log\left(1-t^n\right)} \tag{6}$$

Body part recognition is vital for accurately tracking the human skeleton and detecting specific human joints within a depth image. In this study, we take advantage of the Microsoft offers in the Kinect V2 framework.[25] In this method, a deep random decision forest[25] classifier is used to categorize each pixel as belonging to a specific body part, enabling the recognition of human joints and tracking of human skeletons by using the Kinect V2 camera. This approach can robustly and effectively classify the body parts within a single depth image, enabling precise identification and tracking of specific joints. In particular, our system exploits the body part classification to detect and track hand movements accurately; this means that users can seamlessly interact with the design and intuitively control it through gestures, giving the system high overall performance and usability.

The acceleration and distance-based fall prediction method makes effective use of the ability of the Kinect sensor to track the 25 skeleton joints shown in Fig. 7. One of the fall prediction indexes proposed in this study is the average index of acceleration and distance. This index is based on the nine joint points of the 25 skeleton joints. The nine joints critical for human movement are selected on the basis of the results of the experiments conducted[26] and are listed in Table 1. The vertical distance to joints from the floor and the acceleration of the joints can be calculated by monitoring these joints and their relationships with the extracted floor parameters. The vertical acceleration of these nine points and the distance from the ground are recorded and a threshold to judge whether there is any sign of a fall is set. The fall prediction algorithm determines if the average joint–floor distance d_s is less than 0.7 meters using Eq. (7), and the average vertical acceleration as is less than -7 m/s^2 (i.e., downward) using Eq. (8).

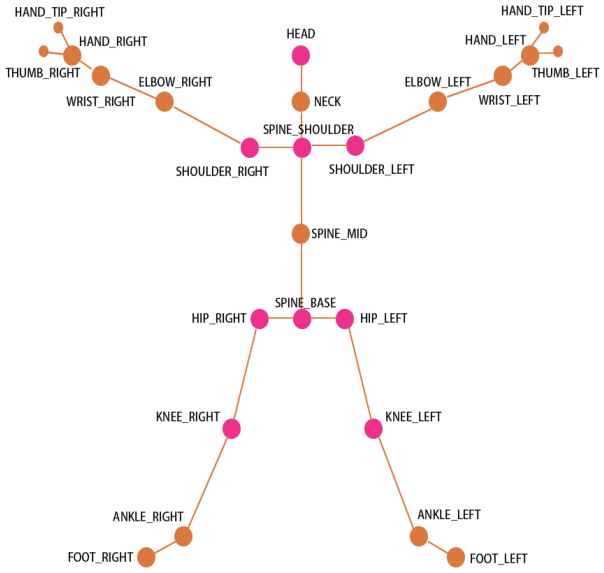

Fig. 7. (Color online) Twenty-five joints tracked by Kinect V2.

Table 1
Joints selected in this study.

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
Head	Spine Shoulder	Shoulder Left	Shoulder Right	Spine Base	Hip Left	Hip Right	Knee Left	Knee Right

$$\frac{1}{9}\sum_{s}^{9} d_s < 0.7 \tag{7}$$

$$\frac{1}{9}\sum_{s}^{9} a_s < -7 \tag{8}$$

In our system, the algorithm predicts a fall if the vertical distance and acceleration surpass their respective thresholds. This prediction is crucial as it triggers an immediate alert, such as visual or auditory notifications or an automated message to a caregiver or medical service. The entity receiving the signal can then initiate appropriate actions to mitigate the risk, such as providing physical assistance, deploying safety measures such as fall-cushioning mats, or preparing for immediate medical attention. Our fall detection approach relies on acceleration and distance-based metrics assessed using a Kinect sensor. The sensor's ability to identify critical joints improves the fall detection accuracy, thereby enhancing the response time. Giving

caregivers or monitoring services a heads-up before the actual fall occurs gives them a window, albeit small, to act in a way that could potentially minimize injury or expedite aid in various environments.

4. Experimental Results

The experimental environment of the program comprised a Kinect V2 RGB-D camera, which was used for data collection, and an alarm message was output to the Kinect's communication software. Kinect V2 captures color and depth images at a resolution of 512 × 424 pixels and a frame rate of 30 fps. The Kinect V2 camera's body part recognition feature proved instrumental in our experiments. It successfully detected 25 joints in the human body, enabling precise tracking and analysis. Moreover, the camera's ability to detect multiple individuals in a scene (up to six people) further enhances its usefulness in various real-world scenarios. The detection range of the Kinect V2 camera is 0.5–4.5 m, enabling effective monitoring and detection of human activity over a sufficient distance for the system to be effective in most indoor environments. The experimental results validated the effectiveness of our approach to providing real-time monitoring and timely alerts.

The depth images used as input are captured by the Kinect V2 sensor. Figure 8 displays a captured depth image of the scene. As shown in Fig. 8, the original depth image is an image with a size of 512 × 424 pixels, which is converted into a disparity image using Eq. (1), and after the accumulation process, the obtained v-disparity image has a size of 256 × 424 pixels, as shown in Fig. 9. The dotted line in the figure indicates the portion of the ground in the original image. The slanted line, seen in Fig. 10, signifies the floor in the experimental scene after Hough transform. To accurately interpret the scene, it is necessary to extract this line. Upon applying the Hough transform, we segment the slanted line depicting the scene's floor, as illustrated in Fig.10. Figure 11 shows the marked floor in the depth image. Furthermore, pixels representing the floor transform and the RANSAC algorithm are employed to extract the floor parameters, thereby enabling us to secure the floor features.

The experiments were conducted in an indoor environment under various light conditions. Two volunteers aged 24 years participated in the tests. They performed various daily activities, including walking, sitting, crouching, and simulating falls in different scenarios. Each fall scenario was repeated at least five times, and the system issued a fall alarm when it predicted a fall would occur. During 10 instances of walking, the system issued no fall alarms. In 20 cases of sitting, the system gave one fall alarm. Similarly, in 20 samples of crouching, the system issued two fall alarms. In 30 simulated falls, the system accurately detected and issued a fall alarm in 24 cases, as listed in Table 2. On average, the fall alarms were given 0.3167 s before the fall event (i.e., the person impacting the floor); the earliest warning occurred 0.5 s before the simulated fall event. The system had high accuracy for distinguishing between walking and falls; however, activities involving downward motion, such as sitting and crouching, occasionally triggered

Fig. 8. Depth image.

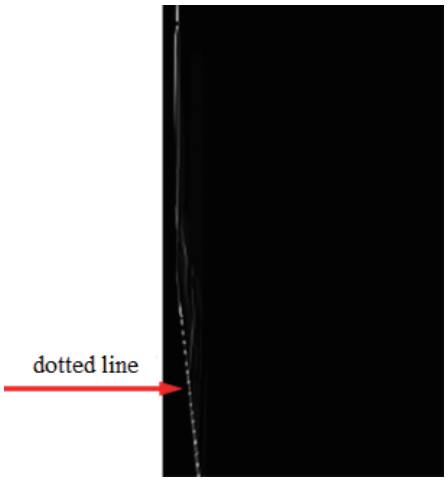

Fig. 9. (Color online) v-disparity image.

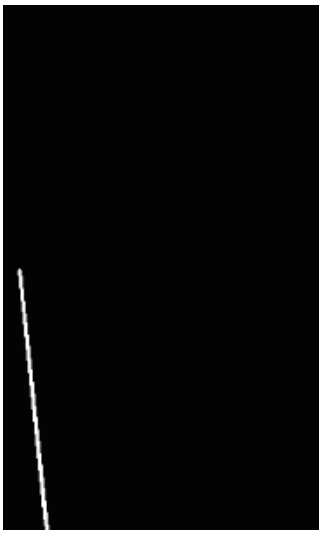

Fig. 10. Hough transform conversion of the slanted line image in Fig. 9.

false fall alarms. Moreover, the system failed to predict some simulated falls. Nevertheless, the accuracy of the expected fall alarms was an impressive 80%. These results indicate the effectiveness of our system in detecting and predicting falls promptly. Moreover, its ability to accurately differentiate between walking and falls suggests that it would be reliable and valid in practical applications. However, the system's detection capabilities during downward motion should be refined to minimize false alarms. Overall, our experimental results demonstrate that the system shows promising performance and the potential to improve safety and well-being in various environments where fall detection is crucial.

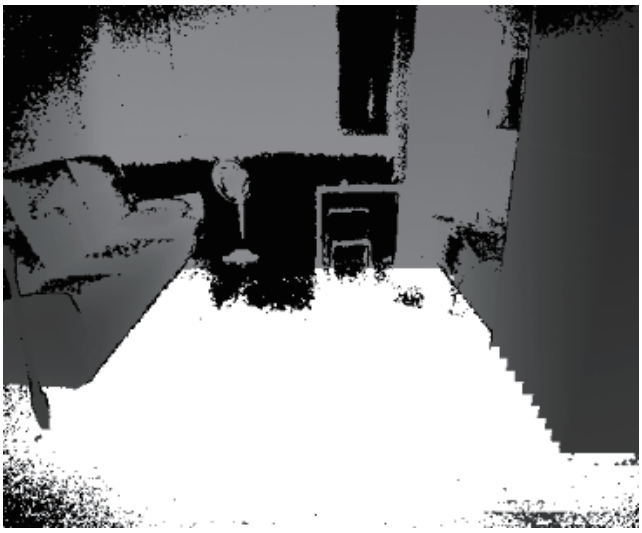

Fig. 11. Floor marked in depth (white area).

Table 2
Experimental results.

	No. of simulations	No. of fall alarms	No. of no alarms	Accuracy (%)
Walk	10	0	10	100
Sit	20	1	19	95
Crouch	20	4	16	80
Fall	30	24	6	80

5. Conclusions

We presented an innovative fall prediction system using the Kinect V2 sensor and image-processing techniques to detect and anticipate falls in real time. The system successfully distinguishes between daily activities and falls and can issue fall alarms up to half a second before a person contacts the ground. The system's primary potential lies in aiding caregivers to promptly assist and minimize fall-related injuries. While the current model achieves an 80% rate, our future goal is to refine the algorithm and harness advanced machine-learning techniques, including deep-learning models, to improve its precision and robustness. Conducting extensive experiments in diverse environments with larger sample sizes would enable a more comprehensive system evaluation, enriching the insight into its reliability and generalizability. This will further facilitate the fine-tuning and optimization of the system parameters.

As part of future work, integrating our fall prediction system with communication software or alert mechanisms would ensure that caregivers or emergency services receive fall alerts promptly, facilitating timely intervention and assistance. Furthermore, augmenting the system with additional sensors, such as inertial sensors or wearable devices, may boost the accuracy and reliability of fall prediction by providing a holistic view of body movement.

Gathering the results of user studies and feedback from individuals at risk of falls and from caregivers will be crucial to further system refinement. Evaluating the system's usability, effectiveness, and user satisfaction within the target demographic will be integral to its improvement. In conclusion, our fall prediction system shows significant potential in mitigating fall-related injuries and ensuring prompt assistance after a fall. The direction of future work includes improving system accuracy, conducting expanded experiments, integrating communication mechanisms, exploring additional sensor modalities, and incorporating user feedback. These efforts are aimed at enhancing the system's performance, usability, and overall effectiveness, thereby contributing to the safety and well-being of at-risk individuals.

References

1 G. Wu: J. Biomech. **11** (2020) 1497. https://doi.org/10.1016/S0021-9290(00)00117-2
2 A. K. Bourke and G. M. Lyons: Med. Eng. Phys. **1** (2008) 84. https://doi.org/10.1016/j.medengphy.2006.12.001
3 A. K. Bourke, J. V. O'Brien, and G. M. Lyons: Gait & Posture **2** (2007) 194. https://doi.org/10.1016/j.gaitpost.2006.09.012
4 J. Qu, C. Wu, Q. Li, T. Wang, and A. H. Soliman: Sens. Mater. **4** (2020) 1209. https://doi.org/10.18494/SAM.2020.2527
5 T. Alanazi, K. Babutain, and G. Muhammad: Appl. Sci. **13** (2023) 6916. https://doi.org/10.3390/app13126916
6 B. M. V. Guerra, S. Ramat, G. Beltrami, and M. Schmid: Front. Bioeng. Biotechnol. **8** (2020) 415. https://doi.org/10.3389/fbioe.2020.00415
7 C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau: IEEE Trans. Circuits Syst. Video Technol. **5** (2011) 611. https://doi.org/10.1109/TCSVT.2011.2129370
8 R. Cucchiara, A. Prati, and R. Vezzani: Expert Syst. **5** (2007) 334. https://doi.org/10.1111/j.1468-0394.2007.00438.x
9 E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, and J. Meunier: IEEE Trans. Inf. Technol. Biomed. **2** (2011) 290. https://doi.org/10.1109/TITB.2010.2087385
10 E. E. Stone and M. Skubic: IEEE J. Biomed. Health Inform. **1** (2015) 290. https://doi.org/10.1109/JBHI.2014.2312180
11 G. Mastorakis and D. Makris: J. Real-Time Image Process. **9** (2014) 635. https://doi.org/10.1007/s11554-012-0246-9
12 Z. A. Mundher and J. Zhong: Int. J. Mater. Mech. Manuf. **2** (2014) 133. https://doi.org/10.7763/IJMMM.2014.V2.115
13 B. Kwolek and M. Kepski: Comput. Methods Programs Biomed. **3** (2014) 489. https://doi.org/10.1016/j.cmpb.2014.09.005
14 Z. Chen, X. Song, Y. Zhang, B. Wei, Y. Liu, Y. Zhao, K. Wang, and S. Shu: Sens. Mater. **3** (2022) 1241. https://doi.org/10.18494/SAM3633
15 N. S. Pai, P. X. Chen, P. Y. Chen, and Z. W. Wang: Sens. Mater. **5** (2022) 1971. https://doi.org/10.18494/SAM3734
16 G. Mastorakis and D. Makris: Sensors **10** (2018) 3363. https://doi.org/10.3390/s18103363
17 N. Lu, Y. Wu, L. Feng, and J. Song: IEEE J. Biomed. Health Inform. **1** (2019) 314. https://doi.org/10.1109/JBHI.2018.2808281
18 J. Shi, D. Chen, and M. Wang: Sensors **17** (2020) 4750. https://doi.org/10.3390/s20174750
19 C. H. Chen, C. M. Kuo, C. Y. Chen, and J. H. Dai: J. Vib. Control **11** (2013) 1603. https://doi.org/10.1177/1077546312449645
20 J. S. Sheu and C. Y. Han: Advan. Technol. Inno. **5** (2019) 10. https://doi.org/10.46604/aiti.2020.4284
21 P. T. Wang, S. Y. Lin, and J. S. Sheu: Advan. Technol. Inno. **4** (2021) 213. https://doi.org/10.46604/aiti.2021.7192
22 P. T. Wang, J. S. Sheu, and J. H. Lai: Sens. Mater. **2** (2022) 563. https://doi.org/10.18494/SAM3616
23 R. Labayrade, D. Aubert, and J. P. Tarel: IEEE Intelligent Vehicle Symp. (2002) 646. https://doi.org/10.1109/IVS.2002.1188024
24 M. A. Fischler and R. C. Bolles: Commun. ACM **6** (1987) 381. https://dl.acm.org/doi/10.1145/358669.358692
25 J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake: CVPR **2011** (2011) 1297. https://doi.org/10.1109/CVPR.2011.5995316
26 S. K. Yadav, K. Tiwari, H. M. Pandey, and S. A. Akbar: Soft Comput. **26** (2022) 877. https://doi.org/10.1007/s00500-021-06238-7

About the Authors

Po-Tong Wang earned his M.S. degree in Computer Science from National Taipei University of Education (NTUE), Taiwan, in 2015. He later obtained his Ph.D. degree from the Bio-Industrial Mechatronics Engineering Department, National Taiwan University (NTU). He is an assistant professor in the Department of Electrical Engineering, Lunghwa University of Science and Technology, Taiwan. His primary research interests encompass a range of fields, including image sensors, color space transformation algorithms, PLCopen control for advanced industrial automation, and applications of industrial image vision and automation in Taiwan. (neojwang@gm.lhu.edu.tw)

Yu-Jen Chen received her B.S. degree from the Department of Computer Science, National Taipei University of Education, Taiwan. Her main interests are in image processing, sensor applications, and embedded systems. (aaronchen8559@gmail.com)

Jia-Shing Sheu received his M.S. and Ph.D. degrees from the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, in 1995 and 2002, respectively. He is currently a professor in the Department of Computer Science, National Taipei University of Education, Taipei, Taiwan. His research interests include pattern recognition and image processing, with a particular focus on real-time face recognition, and embedded systems. (jiashing@tea.ntue.edu.tw)