Development of an intrinsic health risk prediction model for camera-based monitoring of older adults living alone

Sekyoung Youm (skyoum@dgu.edu)
Dongguk University
Minji Kim
Dongguk University
Songiee Hong
Dongguk University

Article

Keywords: behavior monitoring, action detection, intrinsic risk prediction, deep learning

Posted Date: March 25th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1439011/v1

Abstract

Due to the recent COVID-19 pandemic, drastic changes in the social service system for the care and safety of vulnerable social groups are urgently required in South Korea. It is interesting whether welfare technology can temporarily replace or supplement traditional care services, originally performed by trained human resources and further can be applied in a closed and non-contact living space. Therefore, efforts are being made to converge various intelligent information technologies such as artificial intelligence, robots, and IoT in care services. Considering the limitations of the "National Emergency Notification System," which was first implemented for older adults living alone in Korea, we performed to establish a monitoring system that can measure health-threatening behaviors, social isolation, and self-neglect of older adults living alone using RGB video data and the latest deep learning methodologies. We would like to suggest future applicability of welfare technology, which can aid in active health care management instead of just simple monitoring. Such a welfare technology can provide a new alternative form of social service to increasing aging population in single-person households and to overcome changes in elder care environment

Introduction

According to Korea´s statistics on older adults in 2021, older adults comprise 16.5 % of the total population in South Korea. Among them, community older adults living alone, account for 35.1 %, which is expected to continue to rise in the future[1]. According to the statistical data of 2019, as the number of older adults increased, the proportion of de-pressed patients belonging to this category became 40.4 % of the total population, and this proportion is continuously increasing. Depression in the older adult has a negative impact on the overall life of the older adult and may lead to death depending on the degree of depression and prolonged depression[2,3,18]. Among the many factors that cause depression in the older adult, there are social isolation and self-neglect[4,19]. Recently, 'social distancing'[15,16,17], which advises people to stay at home due to the COVID-19 pandemic, has an indirect effect on the onset of depression in the older adult[5]. A life of isolation due to a decrease in intimacy with neighbors and a decrease in interpersonal relationships due to 'social distancing' can lead to depression[19]. the rate of Self-neglect, which means that the older adults do not take care of themselves and are inactive, is increasing every year[20]. Self-neglect also has a significant correlation with depression[4]. In addition, the older adults show irregular nutritional intake due to various factors such as economic status and loss of appetite, and irregular nutritional intake increases nutritional risk factors[21,22]. Such irregular eating habits do not cause health deterioration in a short term, but they can worsen health in a long term. Social isolation, self-neglect, and abnormal nutritional habits can be discovered through observation of daily life patterns. it is mainly observed using the activities of daily living(ADL) assessment. ADL refer to the basic activities that an individual needs to perform in order to function independently in daily life[10]. A correlation between the ADLs and the cognitive dysfunctions has been reported[11]. Since ADLs are used as an indicator to measure the severity of dementia[24] and to determine the effectiveness of dementia treatment agents[12], ADLs can provide specific information on the ability of older adults to manage their daily life

independently[13]. Currently, the ADLs are either conducted with the older adults directly or through observations and interviews with their family members and nurses[13,14]. However, the reliability of such assessments is questionable as they may be affected by subjective factors[12].

In recent times, the fourth industrial revolution is garnering considerable attention, and the technology development to realize it has accelerated. Consequently, numerous studies are being conducted to solve current societal problems via the convergence of the fourth industrial revolution technology[6]. Technologies in monitoring older adults have been proposed as a feasible method to aid in decision making and reduce caregivers' burden[7]. The focus of monitoring technology is on identifying older adults' emergency situations[8,9]. the current monitoring system is unable to detect common lifestyles of older adults, such as inactivity and self-neglect. A major challenge of current technologies can not accurately monitor daily life, which involves uncertain movements.

Therefore, this study was performed to design a monitoring system that can automatically assess social isolation, inactivity, unhealthy behaviors, etc., rather than solely monitoring for life-threatening emergencies currently detected. Furthermore, instead of simple monitoring, the developed system can be applied to the data of older adults' actual daily lives. Therefore, set up two stages of the study procedure: First, we attempt to advance the camera-based technology of monitoring older adults' daily lives; Second, we plan to apply this advanced skill to the real settings of older adults who are institutionalized at nursing homes in Kang Hwa, South Korea. This article focused on the first stage of our study. We expect this camera-based monitoring system to improve the quality of care services in health care settings.

Methods

Materials.

In this study, we used the ETRI-Activity3D dataset by filming the lives of older adults from a robot's point of view to solve the problems of an aging society and to learn the behavior of older adults[32]. The data was collected in a 102m^2 apartment setting reflecting a real-home environment, and it was constructed to elicit similar behaviors of older adults as accurately as possible. Figure 1 shows a typical example of the video data. The actual data used from the ETRI-Activity3D data are shown in Table 1. ETRI-Activity3D was approved on April 20, 2021 by submitting an agreement on the ETRI AI Nanum Website(https://nanum.etri.re.kr/share/dhkim008/robot_environment2?lang=en_KR).

Table 1

Definition of ETRI-Activity3D data

Items	Content
Total number of samples	5,339(Train 4091 / Test 1248)
Number of behavior classes	13
Number of people filmed	20 people (10 older men, 10 older women)
Filming environment	102m^2 apartment living environment
Filming location	Bathroom, kitchen, living room, etc
Used data format	RGB videos
FPS	25

There was a total of 55 classes in the ETRI-Activity3D data set. However, in accordance with the behavior classes for intrinsic risk outlined, the categories were defined as (1) behaviors that can be used to assess the ability to perform ADLs as a basic daily routine; (2) behaviors that may be unhealthy to health in the long term; and (3) behaviors that can indicate the social relationships of the older adults. Therefore, we define a total of 13 classes, and these classes satisfying respective conditions were presented in Table 2. Several behavior classes were merged and redefined into a single class. The behavior of using a vacuum cleaner and that of cleaning the floor while bending forward were combined and defined as “Cleaning the room.” Similarly, reading a book and reading a newspaper were combined and defined as “Reading.” The action of making or receiving a call and the behavior of operating a smartphone were also combined and de-fined as “Using a phone.”

Table 2

Definition of behavior classes used by the developed system

Class criteria	Behavior Class	Total Number of Data
(1) Behaviors that can be used to assess the ability to perform ADL as a basic daily routine	Using a gas stove	266
	Cleaning the room	398
	Cleaning the furniture	286
	Hanging laundry	392
	Reading	558
	Using a remote	324
	Lying	378
(2) Behaviors that may be unhealthy to health in the long term	Eating	507
	Taking medicines	580
	Drinking	376
	Smoking	362
(3) Behaviors that can indicate the social relationships of the older adults	Talking	330
	Using a phone	582

Action recognition

Action recognition technology is essential to monitor older adults' daily lives. since it has been widely applied to video information search, daily life security, and CCTV surveillance[23]. This technology is divided into two types of actions: action classification and action detection[24]. Action Classification implies classifying the type of action a person is performing in a video. Consequently, the video data used as the input should consist of an image of an action. In contrast, action detection detects which action is taken at a certain point in a video that was not cut, based on a specific class criterion[23,24]. Since most of videos collected in a real life are unedited videos that include multiple actions rather than videos that only include a specific action, action detection is essential for recognizing the specific target actions of a person in the actual video[24,25].

The methods of action recognition can be broadly divided into the following: (i) a method that uses RGB image data without changes[26,27,34], and (ii) a method that detects an action using the skeleton

coordinates of human body derived from the RGB image data[28,29]. However, when RGB images are used, information other than actions such as background and color may be also sensitive to action detection[19,21,22,34]. Therefore, if action recognition is performed using the skeleton coordinates of human body in RGB images, then it is possible to overcome the limitations by solely obtaining motion data without being affected by the background or lighting[30,31]. For this reason, Posec3d[28], which performs skeleton-based action recognition, was used in this study.

Posec3d can be divided into two main segments: a pose estimator and a behavior detector. In the pose estimator, a video, entered as the input data, is sliced into images at a certain frame per second, and human skeleton coordinates are derived from each image. In this paper, in the pose estimator stage, the human pose was estimated using the Top-Down method, which has an advantage over the Bottom-Up method in terms of accuracy[36]. Therefore, the human body was first detected and then the human skeleton coordinates were derived. First, the object detection algorithm Faster Region-based Convolutional Network (Faster CNN)[37] was used to detect a person, and the pose estimation algorithm High Resolution Network (HRNet)[35] was used to extract the human skeleton coordinates of the detected person. After converting each extracted human skeleton coordinates into a 2D heatmap, it is stacked according to time flow to construct a 3D heatmap. The behavior detector uses the 3D ResNet-based 3D convolutional neural network (CNN) as the input for a 3D heatmap to detect and identify the actions performed by a person in a video. Figure 2 shows the framework of Posec3d.

Development of and ADL Monitoring System

Framework

The study framework (Figure 3) allows older adults to record their behaviors using Posec3d which is designed to evaluate older adults' ADLs detect the intrinsic risks. Posec3d enables state-of-the-art recording in the field of action recognition using human skeleton coordinates[33].

In this study, we evaluated different behaviors of older adults, such as social isolation, self-neglect, and long-term unhealthy behavior, these behaviors elucidated in this study are: (1) basic behaviors that can evaluate the personal capacity of ADLs, (2) behaviors that can be unhealthy to long-term health, and (3) behaviors that can identify the social relationships of older adults.

Data Processing

All the performance processes were carried out using Python. Figure 4 shows a series of steps used for preprocessing the data. These processes created an annotation source of learning in the Posec3d algorithm.

The training data and test data were divided into a ratio of 8:2 based on the number of participants and were set as the training dataset (8 males, 8 females) and test dataset (2 males, 2 females). The ETRI-Activity3D dataset was a video filmed at 25 fps and sliced to 25 fps using ffmpeg to create several images.

Annotation was a collection of information on the data to be used in the learning. The annotation contained information for each video as shown in Table 3. This information was stored in a dictionary format, and the final annotation was in the form of a list containing the annotation of each video. Here, “Frame_dir” was used to distinguish the name of the video, and “Img_shape” and “Original_shape” was 1080 width and 1920 height (1080, 1920). Because each video had a different length, the total number of frames also varies.

Table 3

Annotation definitions and examples used in the system

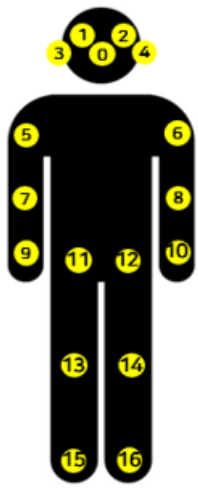
Item	Explanation	Data Format	Explanation on data format
Frame_dir	Name of video	String	Ex) A001_P001_G001_C004
Img_shape	Size of image data	tuple	(height, width) = (1080, 1920)
Original_shape	Original size of video	Tuple	(height, width) = (1080, 1920)
Total_frames	Total number of video frames	Integer	Ex) 456
Keypoint	Skeleton’s x,y coordinates for people inside each frame of the video	Array	[N(number of people), T(number of frames), K(number of keypoint = 17), 2(x,y coordinates)]
Keypoint_score	Confidence value of the skeleton value of the person in each frame in the video	Array	[N(number of people), T(number of frames), K(number of keypoints = 17)]
Label	Class of the video	integer	Ex) 0

The keypoints were extracted using HRnet[35], and then a total of 17 skeleton coordinates were used in Table 4. The order of each skeleton coordinate was specified in Table 4. Here, “Keypoint_score” indicated the confidence value of each keypoint. Therefore, if the confidence value was high, the human skeleton coordinate could be detected with a high accuracy. Furthermore, “Label” represented the information on

the behavior being filmed in the video and showed that each behavior can be expressed using numbers (Table 4).

Table 4

Definitions of the keypoints, used for pose extraction, and behavior class number

	Keypoint		Behavior Class	
	0	nose	0	Eating
	1, 2	eye	1	Taking medicines
	3, 4	ear	2	Drinking
	5, 6	shoulder	3	Using a gas stove
	7, 8	elbow	4	Cleaning the room
	9, 10	wrist	5	Cleaning the furniture
	11, 12	hip	6	Hanging laundry
	13, 14	knee	7	Using a remote
	15, 16	ankle	8	Reading
			9	Using a phone
			10	Smoking
			11	Talking
			12	Lying

Algorithm for judging the behavior of older adult

Posec3d is an algorithm for action classification during action recognition using skeleton coordinates; that is, it is an algorithm that classifies the entire image into a single class. However, it was difficult to use in its original form because older adults' daily behaviors are too complex to be detected, i.e., what behavior is performed at a specific time on basis of daily routine. Therefore, 90 frames were defined as the time to act, and the problem was mitigated by repeating Posec3d after every five frames. Using this Posec3d algorithm, action detection was performed every five frames to detect what actions older adults are doing, and then the actions and respective corresponding time were coincidently recorded in the database. According to the set number of frames per second, the video was divided into images data. For example, based on 25 fps, five images correspond to 0.25 s, and 250 images corresponded to 10 s. Therefore, time information was derived in this way, which allowed us to record respective start time and end time of older adult's activities, configure a database, and derive information on what activities the older adult did at what time through the database. Thus, using this process, his/her repeated daily behaviors can be identified and implemented as the baseline data for ADLs through the stored database.

Experiment

In this experiment, we employed the Ubuntu 18.04 LTS operating system for machine learning using two RTX3090. The total batch size was set to 64, the learning rate was set to 0.025–1000 epochs, and the stochastic gradient descent was used as the optimizer function. In this case, the loss function is cross entropy. The model calculation of such leaning process was verified using the test annotation every 10 epochs, and the final model yielded the highest accuracy of older adults' target behaviors. Specifically, among the 100 validations out of 1000 epochs, the 940-epoch model was selected as the best performance shown in the Figure 5.

<Figure 5 about here>

Results

The evaluation of test dataset

To evaluate the accuracy of machine learning and deep learning models, we utilized the performance indicators such as accuracy, precision, recall, and F1-score. Empirically, we obtained an excellent behavioral performance showing that all indicators were accepted as the 98 % accuracy, 98 % precision, 99 % recall, and 98 % F1-score.

In Figure 6 (a), the actual label value and predicted label value were expressed as a confusion matrix. Since the number of labels for each class was not the same, as shown in Figure 6 (b), we normalized the number of each class. The x- and y-axes in Figure 6 (a & b) represent the predicted and actual labels, respectively. Table 5 presented the performance indicators for each class.

Table 5

Performance indicators for each behavior class of the system

The application of real data The model was applied to actual videos to verify whether the actions shown in them were properly predicted or not. As shown in Figure 7, the model was applied in older adults' actual residential space to test and confirm that their actions were recorded in the database. In addition, it was applied to various existing videos to verify its workability in a real environment. Experimental data for verification has been approved by the institutional review board of Dongguk University(DUIRB-202106-15). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects and/or legal guardians that their information/images will be included in an online open-access publication and paper. And we confirmed that the informed consent was obtained from all the subjects (for participation). The data shown in Figure 7 (a), (b) is data collected in a real environment, and permission was obtained from the person in the video to use the data. The data shown in Figure 7 (c), (d) is data published on YouTube[38]. In the case of the upper data shown in Figure 7, While the Data (a) and (b) were possible to be transformed as a behavioral database because the actual shooting time was 15:00, the data (c) and (d) were not because the actual time was unknown from the news. For this reason, the database was built to display changes in behavior in seconds by

Class	Action	Accuracy	Precision	Recall	F1-Score
0	Eating	89%	96%	89%	93%
1	Taking medicines	99%	100%	99%	100%
2	Drinking	99%	99%	99%	99%
3	Using a gas stove	100%	98%	100%	99%
4	Cleaning the room	100%	100%	100%	100%
5	Cleaning the furniture	99%	100%	99%	99%
6	Hanging laundry	100%	100%	100%	100%
7	Using a remote	99%	100%	99%	99%
8	Reading	99%	94%	99%	96%
9	Using a phone	99%	98%	99%	99%
10	Smoking	98%	94%	98%	96%
11	Talking	100%	98%	100%	99%
12	Lying	100%	100%	100%	100%

quickly replaying the human videos in the news.

In this algorithm, false detection often occurred among older adults' behaviors appeared to be similar as human skeleton coordinates. For example, misdetection of "Using a phone" from "Using a remote" occurred because the behavior of reaching forward to use a cell phone was similar with reaching out to grasp a TV remote controller.

Conclusions

We used a Posec3d-based algorithm that performs deep learning-based action recognition to develop an older adult monitoring system which can monitor older adults living alone in a residential setting. Such a monitoring system organized and stored actions in a database depending on the start time and end time of certain actions was proposed. In particular, older adults' behaviors that can cause social isolation, self-neglect, and health deterioration were defined as intrinsic risks in this study. Accordingly, three behavior categories were defined: (1) behaviors that can assess the ability to perform ADL in daily life, (2) behavior that can be unhealthy to health in the long run, and (3) behaviors that indicate the social relationship of older adults. The ETRI-Activity3D data, which includes videos of older adults, was used. Compared to other behavioral data, this dataset shows distinctive physical features of older adults, such as a curved spine different from that in young adults. In this way, we could perform the specific monitoring customized for older adults.

Various older adult data could not be used because of storage limitations. To overcome the lack of diversity in the study data, we will apply this study technology into older adults' real settings in the next stage of this research, including personal houses and institutional settings like nursing homes. This study method was that the human skeleton data, which only contains motion information. Therefore, limitation is similar actions are not properly detected. Therefore, in future studies, action recognition should be

performed using the RGB image data and skeleton data in multimodal form, and more accurate action recognition studies should be conducted.

Declarations

Author Contributions

M.K. was responsible for methodology, analysis, and writing—original draft preparation. S.Y. was responsible for methodology, review, and editing. S.H. was responsible for the conceptualization and review. All authors have read and agreed to the published version of the manuscript.

Corresponding authors

Co-Correspondence to Sekyoung Youm and Songiee Hong.

Competing interests

The authors declare no competing interests.

Data Availability

RGB Dataset for recognizing the daily behavior of older adults in a robotic environment provided by ETRI was used. The data provided by ETRI have been obtained from 100 persons (50 older adults and 50 general adults), and there are a total of 55 daily behavior classes. When observing the behavior of older adults, 55 types of behavior classes were constructed based on frequent activities, and in this study, 16 classes taken by 10 males and 10 females were used and reconstructed into 13 classes. Anyone can download ETRI-Activity 3D Data after applying for data on the ETRI nanum website (https://nanum.etri.re.kr/share/dhkim008/robot_environment2?lang=en_KR). So we don't have permission to share.

Acknowledgments

This work was supported by the National Research Foundation of Korea(NRF- 2020R1A2C2010471)

The experimental data obtained permission was obtained from the person in the video to use the data, and was reviewed and approved by institutional review board of Dongguk University (DUIRB-202106-15).

References

1. Statistical Office. *2021 Older Adult Statistics.* (2021).
2. Oh, H. Relationship between social capital, depression and quality of life in elderly people participating in physical activity. *The Korean journal of physical education* **53**, 535-547 (2014).

3. Waern, M., Rubenowitz, E. & Wilhelmson, K. Predictors of suicide in the old elderly. *Gerontology* **49**, 328-334 (2003).
4. Dong, X. *et al.* Elder self-neglect and abuse and mortality risk in a community-dwelling population. *Jama* **302**, 517-526 (2009).
5. Chan, A., Malhotra, C., Malhotra, R. & Østbye, T. Living arrangements, social networks and depressive symptoms among older men and women in Singapore. *International Journal of Geriatric Psychiatry* **26**, 630-639 (2011).
6. Choi, S., Kim, C., Kang, Y. & Youm, S. Human behavioral pattern analysis-based anomaly detection system in residential space. *The Journal of Supercomputing* **77**, 9248-9265 (2021).
7. Camp, N. *et al.* Technology used to recognize activities of daily living in community-dwelling older adults. *International Journal of Environmental Research and Public Health* **18**, 163 (2021).
8. Won, J., Kim, C., Choi, S., Youm, S. & Kang, Y. S. TensorFlow object detection api-based pose identification procedure for elderly living alone emergencysituation detection. *The Korean Institute of Information Scientists and Engineers*, 726-728 (2018).
9. Kim, G. & Park, S. Activity Detection from Electricity Consumption and Communication Usage Data for Monitoring Lonely Deaths. *Sensors* **21**, 3016 (2021).
10. Vermeulen, J., Neyens, J. C., van Rossum, E., Spreeuwenberg, M. D. & de Witte, L. P. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatrics* **11**, 1-11 (2011).
11. Gold, D. A. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *Journal of clinical and experimental neuropsychology* **34**, 11-34 (2012)
12. Bavazzano, A. *et al.* Functional evaluation of Alzheimer patients during clinical trials: a review. *Archives of Gerontology and Geriatrics* **26**, 27-32 (1998).
13. Yang, Y. et al. Activities of daily living and dementia. *Dementia and Neurocognitive Disorders* **11**, 29-37 (2012).
14. Jang, J. Comparison of activities of daily living differences with dementia stage. *Journal of the Korea Academia-Industrial cooperation Society* **18**, 557-563 (2017).
15. Morley, J. E. & Vellas, B. *The journal of nutrition, health & aging* 24, 364-365 (2020).
16. Lim, W. S. *et al.* COVID-19 and older people in Asia: Asian Working Group for Sarcopenia calls to action. *Geriatrics & gerontology international* **20**, 547-558 (2020).
17. Kim, J., Kim, Y. & Ha, J. Changes in daily life during the COVID-19 pandemic among south korean older adults with chronic diseases: a qualitative study. *International Journal of Environmental Research and Public Health* **18**, 6781 (2021).
18. Plagg, B., Engl, A., Piccoliori, G. & Eisendle, K. Prolonged social isolation of the elderly during COVID-19: Between benefit and damage. *Archives of gerontology and geriatrics* **89**, 104086 (2020).
19. Kim, M., Eo, Y. & Kim, S. A study of depression in the elderly by individual and community effects. *Health Soc. Welf. Rev* **39**, 192-221 (2019).

20. Chang, S. & Kim, S. The social network typology among elderly living alone in Busan, depression, and self-neglect. *Korean Journal of Gerontological Social Welfare* **72**, 245-273 (2017).
21. Schlenker, E. Nutrition in Aging. 2nd ed.1993, 186-195 (WCB McGraw-Hill, 1993).
22. Solomons, N. W. Nutrition and aging: potentials and problems for research in developing countries. *Nutrition reviews* **50**, 224-229 (1992).
23. Yao, G., Lei, T. & Zhong, J. A review of convolutional-neural-network-based action recognition. *Pattern Recognition Letters* **118**, 14-22 (2019).
24. Moon, J., Kim, H. & Park, J. Trends in Temporal Action Detection in Untrimmed Videos. *Electronics and Telecommunications Trends* **35**, 20-33 (2020).
25. Wu, D., Sharma, N. & Blumenstein, M. in *2017 International Joint Conference on Neural Networks (IJCNN).* 2865-2872 (IEEE).
26. Feichtenhofer, C., Fan, H., Malik, J. & He, K. SlowFast Networks for Video Recognition. *Proceedings of the IEEE/CVF international conference on computer vision.* 6202-6211.
27. Carreira, J. & Zisserman, A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 6299-6308.
28. Duan, H. *et al.* Revisiting skeleton-based action recognition. *arXiv preprint arXiv:2104.13586* (2021).
29. Yan, S., Xiong, Y. & Lin, D. Spatial temporal graph convolutional networks ofr skeleton-based action recognition. *Thirty-second AAAI conference on artificial intelligence.*
30. Vemulapalli, R., Arrate, F. & Chellappa, R. Human action recognition by representing 3D skeleton based action recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 588-595.
31. Du, Y., Wang, W. & Wang, L. Hierarchical recurrent neural network for skeleton based action recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1110-1118.
32. Jang, J. *et al.* ETRI-Activity 3D: A Large-Scale RGB-D Dataset for Robots to Recognize Daily Activities of the Elderly. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).* 10990-10997 (IEEE).
33. "Skeleton based Action Recognition on NTU RGB+D State of the art" paperswithcode "https://paperswithcode.com/sota/skeleton-based-action-recognition-on-ntu-rgbd" (assessed on 7 October 2021)
34. Sun, Z. *et al.* Human action recognition from various data modalities: A review. *arXiv preprint arXiv:2012.11866* (2020).
35. Sun, K., Xiao, B., Liu, D. & Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5693-5703.
36. Dang, Q., Yin, J., Wang, B. & Zheng, W. Deep learning based 2d human pose estimation: A survey. *Tsinghua Science and Technology* **24**, 663-676 (2019).

37. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28** (2015).
38. JTBC News. "I can't even say a few words a day"... older adults living alone 'shade of non-face-to-face' *Youtube* https://www.youtube.com/watch?v=6VzVSW7olWM (2021).

Figures

Figure 1

Sample of the ETRI-Acticity3D dataset

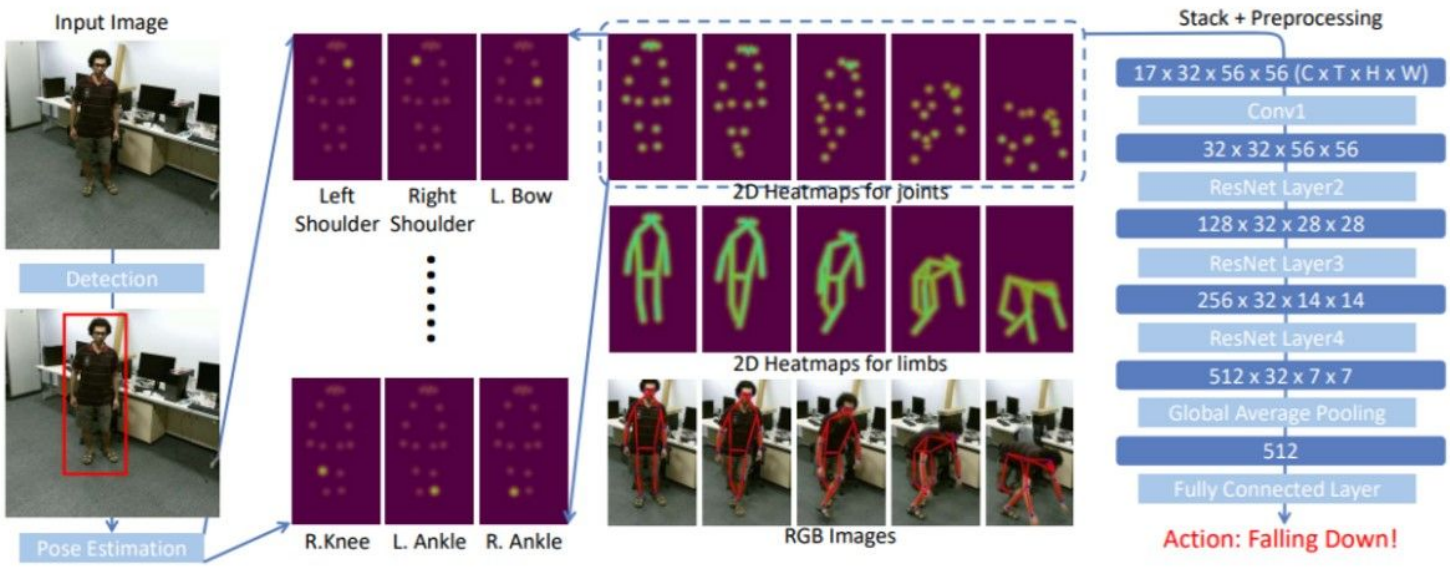


Figure 2

Posec3d framework[28]

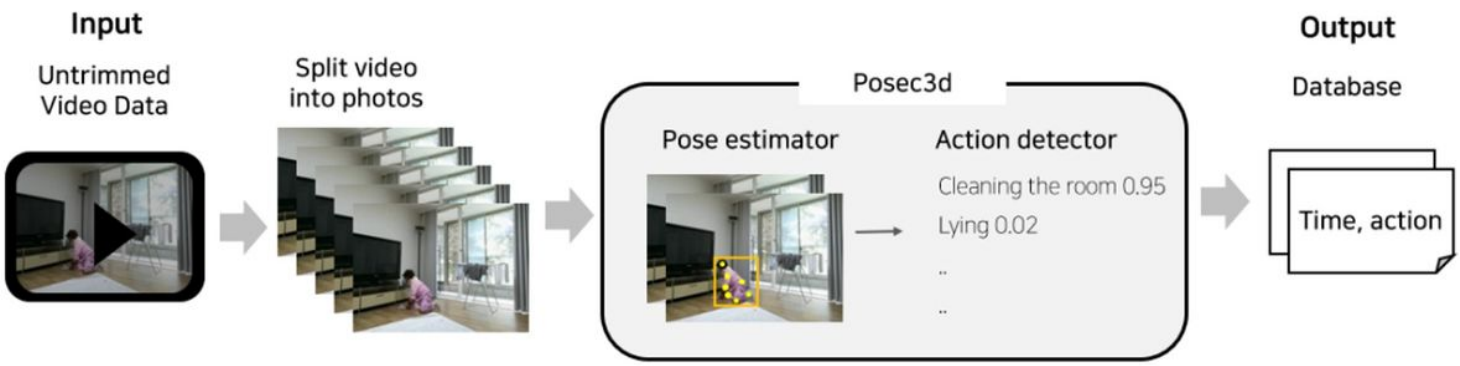


Figure 3

Behavior monitoring system framework

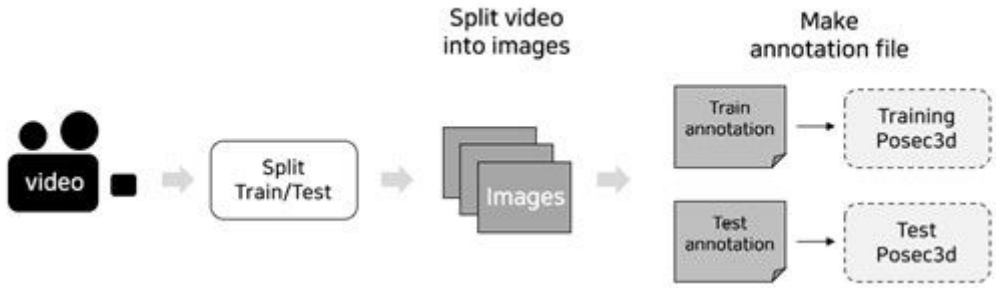


Figure 4

Illustration of the process of creating annotations

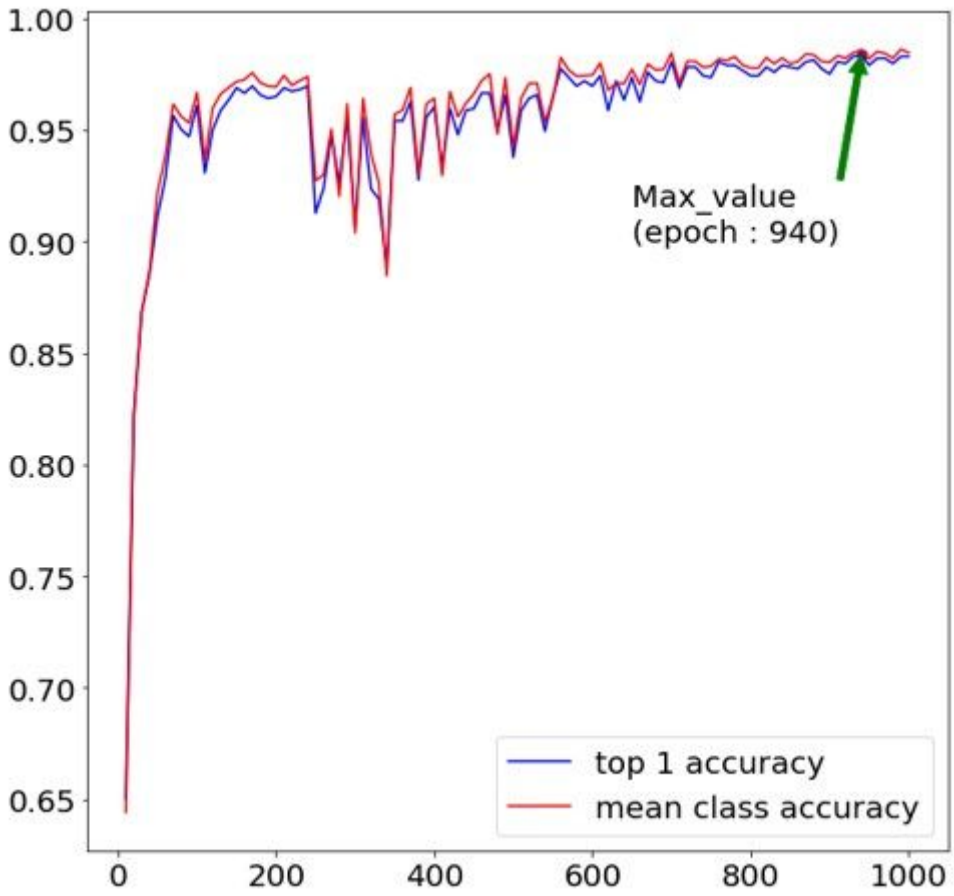


Figure 5

Result of training Posec3d. Based on the highest accuracy, the highest 940 epoch was selected as the final model

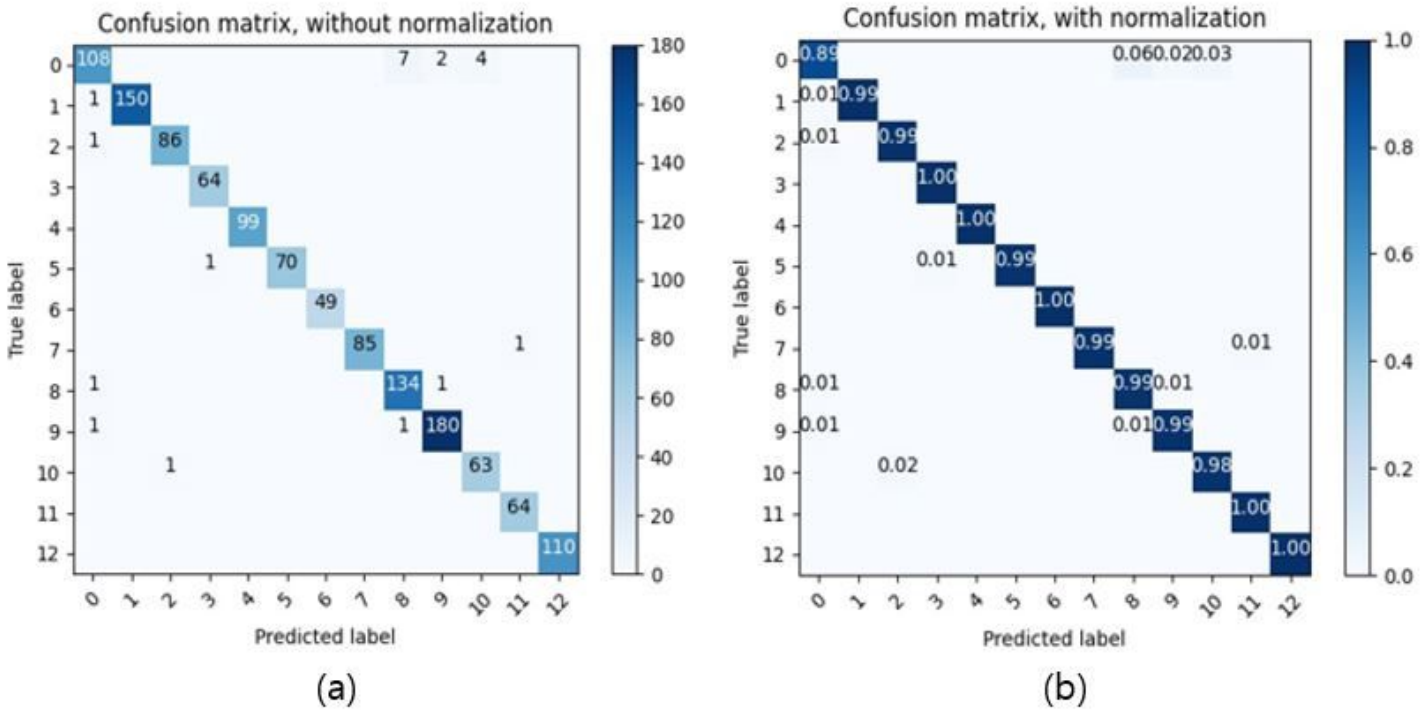


Figure 6

Prediction result confusion matrix of the behavior monitoring system

(a) Non-standard confusion matrix

(b) Standard confusion matrix according to the number of each class

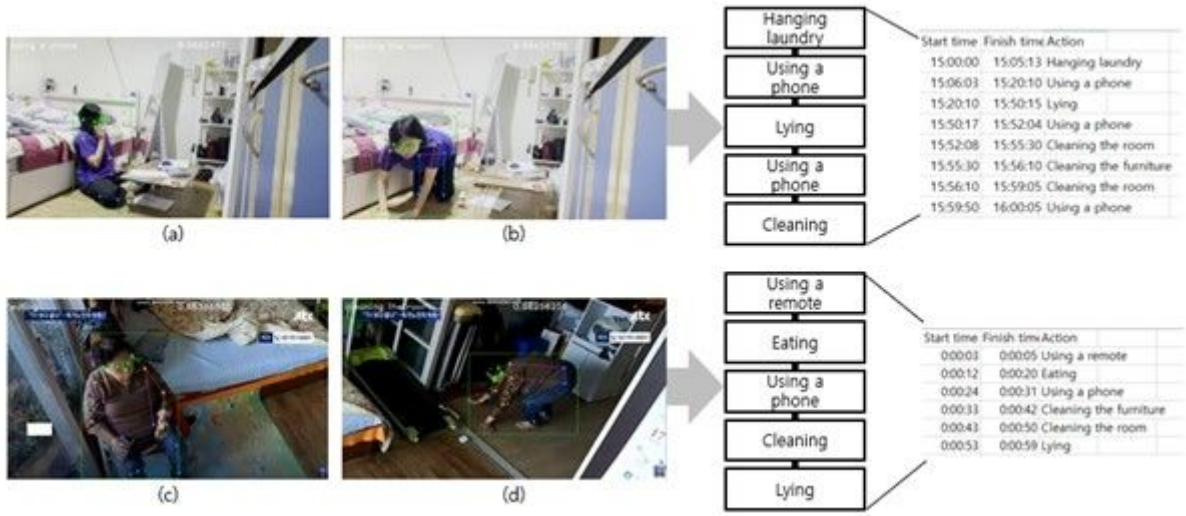


Figure 7

Application of the proposed method to monitor the behavior of an actual older adult

(a) (b) Validation through collected data

(a) (d) Validation using online news data