Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.2204.11

2022, 41(4) 106-116

Low Cost Intelligent Computer Vision based Assistive Technology for Elderly People

Dennis Cheung [a], Peter Joe [a], Muhammad Nadeem [b,*], Saleem Akhtar [c], M. Hassan Aslam [c]

[a] *Department of Electrical, Computer, and Software Engineering, The University of Auckland, New Zealand*

[b] *College of Engineering and Technology, American University of the Middle East, Kuwait*

[c] *Department of Electrical and Computer Engineering, COMSATS University Islamabad, Lahore Campus, Pakistan*

[*] Corresponding author: Muhammad Nadeem, Email: Muhammad.Nadeem@aum.edu.kw

Received: 10 June 2021, Accepted: 25 February 2022, Published: 01 October 2022

KEYWORDS

Computer-Vision

Human Detection

Posture Recognition

Machine Learning

Neural Networks

ABSTRACT

The elderly population (aged above 65) has surpassed the number of children under five years and constitutes 10% of the population of the world, which is expected to reach up to 15% by 2050. A vast majority of the elderly suffer from some kind of chronic disease affecting their cognitive skills impairing their ability to spend a quality life independently. They have to rely on caretaker services for a quality life in a safer environment. However, caretaker services are quite expensive and it is hard for the health care organizations or the families to bear the expenses. 24/7 availability of caretakers is also an issue due to the shortage of workforce, failing to meet the ever-growing demand of the caretakers. This article presents a computer-vision based assistive system for the elderly to tackle this important problem. The system is capable of monitoring the environment as well as the activities of the subject in a standard room environment. It takes the video feed as an input and recognizes the activities of elderly people by using machine learning algorithms. The standalone system autonomously detects and generates real-time alerts for the caretaker, guardian or family member in case of any potential danger through an accompanying smartphone application. It also keeps records of the activities of the subject and generates reports which can be used by health professionals for further analysis and diagnosis. The proposed system was tested in a real-world environment and exhibited promising results for standing, sitting and resting, with a combined average accuracy of 83.41%. The proposed system is standalone, cost effective, and flexible for monitoring the activities of the elderly and is expected to enhance the safety and quality of life of the subject.

1. Introduction

The use of assistive technologies has increased in recent years, due to a rise in the aging population. As per the survey report, the people aged above 65 are 10% of the world population and is expected to increase by 5% till the end of 2050 [1]. The elderly population constitutes 15% of New Zealand and are expected to reach up to 25% by 2050 [2]. Elderly people often suffer from chronic diseases making it difficult for them to survive on their own. Therefore, they need human assistance which cost fortune and compromising on this may pose a threat to the safety as well as the quality of life. Existing solutions to the aforementioned issue are the retirement/old age homes. These homes are purposely built for the elderly taking into account their needs and caretakers are made available around the clock to facilitate them. The facilities and service provided improve the quality of life and provide a safer environment at the same time. However, this solution is very costly and may lead to financial crisis for the family or guardian. The approximate annual cost of a caretaker in New Zealand is around $40,000 [3]. The said amount is quite high and approximately equals to half of the average annual income of a New Zealander [4]. This cost of caretaker is expected to increase in the coming years as per the increase in population above 65 years.

According to a survey, the average age of living alone in 2013 was 62 years, out of which, 28.8% of the elderly live on their own and 51.1% live with their spouse [4]. As per the reports of the World Health Organization (WHO), each year 28-35% of elderly (Aged 64) are involved in falling accidents and this ratio increases to 42% in the age group of 70 and above [5]. As a result of these falling accidents in elderlies, 90% got hip or wrist fractures, whereas 60% got head injuries [6]. Considering these numbers, Retirement/Old age homes has necessary arrangements to handle these falling accidents. These arrangements are limited while living alone at home, which causes a higher ratio of falling accidents and injuries. Moreover, the elderly requires immediate attention in case of any emergency and a delay in providing assistance may result in catastrophe. It is hard to get human assistance on time without the use of technology. The proposed low-cost assistive technology system is capable of monitoring the activities autonomously and providing alerts to the caretaker or healthcare professions via mobile application in case of emergency. The designed solution is very cheap as compared to commercially available assistive technologies which are limited in number. The non-intrusive nature of the system makes it easy to use and does create any discomfort to otherwise cautious elderly. Furthermore, this is a standalone system which was designed using commonly available hardware such as an inexpensive webcam to get a visual feed of the environment and Raspberry Pi to perform computations on vision-based models to identify humans and their activities. We made use of machine learning algorithms to enhance the accuracy of the system.

2. Related Work

Assistive technologies are to facilitate the person with disabilities or to improve the daily functionalities of an individual. Basic modules required for any assistive technology include sensing and decision-making modules [7]. The sensing module is to get the information about the user and environment. These sensing modules are further categorized into three types, (1) sensor-based module, (2) sensor and camera-based hybrid module, and (3) only camera-based module.

Sensor-based modules use accelerometers and gyroscopes to monitor the activities. This type of modules processes the data locally and send processed information to the server for further usage. These sensor-based modules can also be placed at different parts of the body to gather the relevant information. Different parameters such as heartrate, oxygen saturation and blood pressure etc. can be monitored and analysed for further use. On the other hand, accelerometers and gyroscopes are very effective for monitoring activities such as running, walking and lying etc. Accelerometers are good at sudden motion detection, therefore, they are very useful for fall detection.

Many systems deployed for activity monitoring follow the similar architecture: the system is equipped with an accelerometer or gyroscope attached to the body of the subject and from the readings of accelerometer or gyroscope, activities of the person are recognized. Researchers have used accelerometers or gyroscopes to monitor the activities by detecting the local minima and maxima of the sensors' output waveform [8-10]. On the bases of these values, thresholds are set and when the real-time signal passes these thresholds, activities are detected. However, there is a common problem with such sensor-based activity monitoring, that is the similarity between different activity waveforms. For example, the wave pattern for a person sitting is very similar to the person lying. Moreover, the wave pattern for a person lying on bed and falling has also similar patterns. Therefore, fall identification via sensor-based system has higher rate of false detection. Researchers have improved the accuracy of activity monitoring by adding additional parameters and, as a result [10-11], the cost and energy consumption of the device also increases.

Smartphones are the most common devices that people use in their daily life. Moreover, the smartphones are pre-equipped with sensors (such as accelerometers, gyroscopes and camera) and high-end processing unit.

Therefore, their acceptability in assistive healthcare is quite high. Furthermore, if there is a requirement of any other peripheral to use, it can easily be interfaced with the smart phone. Hu et al. [12] used smart phone and 360 omni-lens to build a 3D environment by sending video stream to a server. Later, the server processes the frames to identify the user in the 3D environment. There are some other smartphone applications like Toss 'N' Turn sense, My Sleep APP, Sleep as Android, Runtastic Sleep Better, and iSleep, which use onboard sensors to monitor the sleep patterns of the user [13]. In terms of sensors, there are other dedicated devices like smartwatches, wristbands, and headbands which are being used for healthcare applications. For Instance, OURA, Zeo, WakeMate, UP3, SleepImage, Jawbone Lark, Hexoskin and The Fitbit Charge 2 are some renowned commercially available devices for sleep tracking [14]. These devices are quite expensive and not very comfortable to wear all the time.

To detect the falls in elderly people, there are devices which uses magnetometer along with accelerometer and gyroscope to detect the falls. Moreover, these devices use machine learning algorithms to identify the activity pattern [15]. In these types of devices, initially, a model is trained for the associated activities. Later, by comparing the real time data with trained data using machine learning techniques, efficient and accurate results for activity monitoring can be achieved. These are efficient in term of accuracy, but they are more power consuming as compared to sensor-based devices due to continual comparing with the trained data. When using sensor-based systems or smart phones, power consumption can be reduced by utilizing smartphone only as communication device and sensor can perform the actual sensing. The smartphone logs the sensor data and in case of any emergency sends alters to the selected contacts. Similarly, like sensor-based system, the proposed system allows the user to monitor vital parameters and other information required for healthcare. Later, the said information, sent to the medical centre for further analysis on the vital parameters and other information gathered from sensors. Contrary to elderly's' desire of minimum interference, this device may be intrusive because of many sensors attached to the user's body.

Considering the intrusiveness of sensor-based devices, camera-based technologies are also being used. This technology mainly uses computer vision, machine learning, and deep learning to perform the desired tasks. For instance, to monitor the activities of a user, computer vision techniques are used to determine the posture and action taken by the user, later, by applying machine learning and deep learning techniques activities are recognized in real time [16-17]. To facilitate the researcher community, Google has developed an open-source framework (TensorFlow) to implement deep learning techniques. In [18], researchers used the TensorFlow to detect and identify specific paintings. A wearable camera, the Technaxx Video Sport Sunglasses, captures the video of the painting in real-time. The video feed is compared with already trained model to detect the paintings. Initially, the system has been trained via TensorFlow on desktop using the images feed. The trained data has generated two files, protobuf file and text file. To implement on a real time system, these files are stored in Raspberry Pi. This enables the real time monitoring using camera and Raspberry Pi, installed on a wearable sunglass.

Camera-based solutions are mostly dependent on machine learning algorithm to interpret the visual information. However, there are some other techniques as well to understand the visual information. For instance, researcher used local feature extraction to identify the images. To detect the humans, researchers isolate the distinct colours by using image processing techniques, moreover, to eliminate the noise from the image, they used HSV colour space [19]. Later, a box is placed around the human and changes are being detected if the human moves away from the box in the coming images.

To identify the postures, two techniques are being used most commonly, (1) pressure sensing and (2) camera-based sensing. In first technique, pressure sensors are used to detect the posture of the human body, but this need a lot of sensors to be operatable in a daily life environment. Whereas camera-based sensing performs the same task by implementing fewer cameras in the system [20]. In home environment, to detect the fall and posture of elderly people a common camera has been proposed [21]. To detect the human from the video feed, they applied background subtraction techniques along with projection histogram. For the classification of posture and fall detection, Support Vector Machine (SVM), a machine learning technique is being used. Martin et al. [22] used a 3D camera and Mickrosoft Kinect sensor to detect the posture of person, this system is also developed for residential environment. Bhatia et al. [23] used a multi-view Eigen model to detect the human body movements. Grimm et al. [24] proposed a system based on Bed Aligned Maps (BAMs). The BAMs uses, pressure arrays and a single depth camera to build a complete map. The output accuracy for this method especially for sleep pose is outstanding but this does not consider the motion of the user. Torres et al. classifies six different sleep positions by using a pressure mattress and infrared depth camera [25]. The issue with the said technique is only one fixed camera is used, therefore, alignment problems are not handled as per the definition of Sleep Assessment and Advisory Service (SAAS), UK. Lee et al. [26] explained about a system, which can also classify six sleeping positions. For that, they used Kinect camera hanging over the bed. They used the position of body joints to detect the posture of the user using a parametric approach, but this approach has a major concern that, the user cannot use blanket while sleeping. Evaluation and results are not provided. Huang et al. [27] used both pressure sensors and video images to improve the accuracy researchers. And, Martinez et al. [28] proposed a device to detect sleeping posture and user moments by using BAM descriptor model. The said research is also extended to recognize activities of higher complexity e.g., removing of bed covers [29].

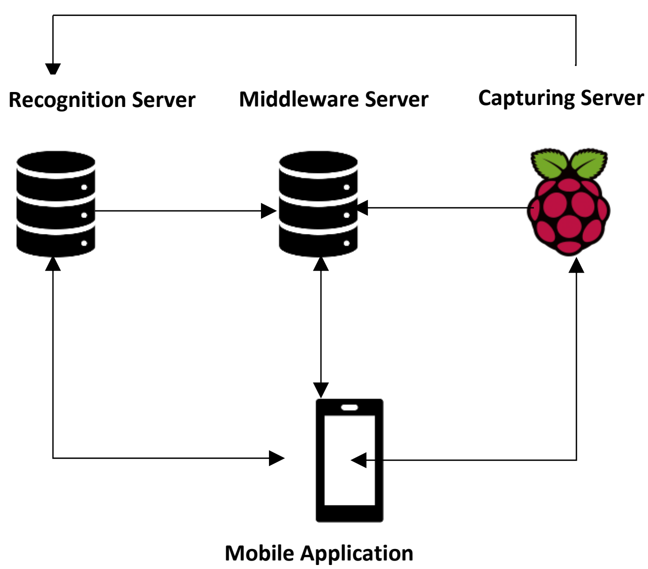


Fig. 1. System architecture

3. System Architecture

The main objective of this research work is to monitor the activities of elderly people by using non-intrusive techniques. For this, a smart and low-cost activity recognition system is proposed in hospice care, which uses visual information as input parameter. The said system is capable of autonomously detecting certain activities of a person in a standard room environment. Fig. 1 shows the conceptual model of the system and also defines the structure of the system.

The front-end of the activity recognition system is an image-capturing device. For capturing the visual information, a 5-mexpixel camera (Logitech C170) is being used, a cost-effective camera with good resolution capabilities. The aforementioned camera is connected with Raspberry Pi 3 B+, which acts as a control unit for the device. The camera provides the video feed to the control unit. The control unit identifies the human presence in the video frame and forwards that particular frame for classification of the activity. The control unit is also equipped with wireless internet module for communication with backend servers. The control unit sends the frame with human presence to the recognition server as Base64 image (Base64 is an encoding scheme used to convert binary data into text format), the recognition server further classifies the activity of the user. Initially, when the system is powered up for the first time, the control unit initiates a communication with middleware server and forward its server URL for future communication. Meanwhile, a JavaScript (JS) and a timed image capturing process is executed simultaneously. The image capturing is performed with the help of a JS library called "node-webcam". This API is capable of detecting the attached camera as well as initiate its capturing facility. After completing the interfacing of camera with the control unit, recognition server is called for image processing. A mobile application is used for controlling different functionalities, e.g. start capturing, stop capturing, and enable/disable image storage option.

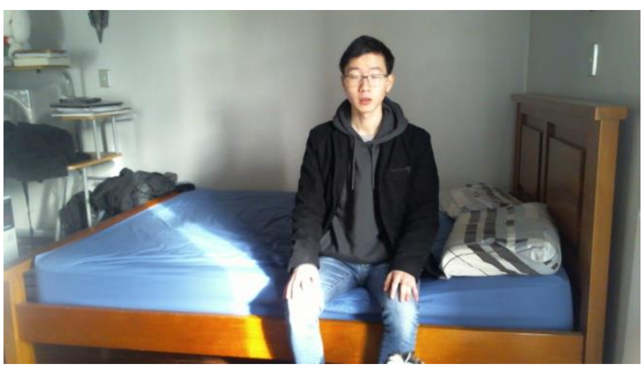
The main functionality of Middleware server is to acts as a communication link between the recognition server and the Raspberry Pi server. The middleware server holds the URLs of the Raspberry Pi server and the recognition server. For consistent communication between various units of the system, the Raspberry Pi server uses dynamic IPs.

On the other hand, the core responsibility of the recognition server is to identify the posture of a human from the input feed. Machine learning algorithms are being executed on the server for classification and identification of the human activities. If the image storage option is enabled via mobile application, the server also stores the image of identified posture with user ID and timestamp for later use. If the mobile application needs to access the database of identified postures, the server also provides the APIs for the mobile application. The recognition server also shares it URL to middleware server at the time of booting up for future communication between the units of the system.

The mobile application is providing a layout to the user to get useful statistics about the person being monitored and to control all the system units. The mobile application is developed using JavaScript framework ‘React Native’. This framework allows easy and quick development of application along with native support. As the application launches, it starts communicating with the middleware server for the URLs of recognition and raspberry server, so the application can communicate with the servers directly. The mobile application can send a request to the recognition server for login and authorization of the user to access the database. Moreover, to change the capturing setting of the camera, it can communicate with the raspberry pi server. Furthermore, the application can show the information stored in the database.

4. Methodology

The posture recognition of human is split into two parts, (1) the detection of human, and (2) the recognition of posture. This methodology is used to avoid processing any redundant information by the recognition server. The image is captured by the camera as shown in Fig. 2(a), then a blue bonding box is being placed on the detected human as shown in Fig. 2(b), and finally the cropped image shown in Fig. 2(c) is sent to recognition server for activity classification.



(a)

(b) (c)

Fig. 2. Posture recognition process, (a) Input image from the webcam, (b) Identification of human detection by blue bounding box, and (c) Cropped image with removal of surroundings

4.1 Detection of Human

First step is to detect the human from the camera feed, followed by the isolation of detected human from the environment. As mentioned earlier and shown in Fig. 2, after detection of human a bounding box is placed and later the image is cropped before sending it to recognition server. Moreover, at this stage, the system also detects the bed in the image, so that, the relevant information can be utilized while classifying the activity specially resting or lying. For human detection, an ImageAI model (a free source python library used for deep learning and computer vision applications) is used along with COCO dataset, a large dataset for Common Objects in Context [30]. The said model has various machine learning algorithms e.g., YOLO, TinyYOLO and RetinaNet, out of which RetinaNet has been utilized for our system. These three algorithms are single stage detector and non-RPN (Region Proposal Network). Therefore, using a large number of candidate object location (approximately 100k) instead of narrowing down the number of candidate object locations to a smaller number. The reason for choosing RetinaNet algorithm for human and bed detection is its highest detection accuracy among all [30]. Contrary to that, it’s the slowest among all, but at the first step its necessary to have higher accuracy rate than efficiency.

4.2 Recognition of Posture

After the detection process is completed, cropped image is forward to the recognition server for posture recognition. Four poses have been aimed for identification i.e., sitting, standing, lying, and resting/sleeping. In the existing literature, there is no complete recognition model available for the mentioned task, therefore, a custom model has been developed and trained by using Image AI. The dataset has been gathered for training the model, for that royalty-free images has been taken. The MPII human pose dataset was also available for the training, but that dataset labeled the images by some actions i.e., Playing board games, reading and fishing etc, instead of general poses i.e., sitting, standing and lying. To train the customized model approximately 500 images per posture were used. Moreover, for testing of each posture a dataset of 150 images per posture was used. As mentioned earlier, initially human is detected from the image to avoid redundant processing, later, the cropped images are passed to the customized model for activity recognition.

There are multiple algorithms available to train the model. However, the ImageAI offers four Convolutional Neural Networks (CNNs), i.e., ResNet50, DenseNet121, SqueezeNet and InceptionV3. Out of these available options, ResNet50 is used due to its time efficiency and higher accuracy [30]. Therefore, initially RetinaNet is used for human detection and then cropped image is passed to ResNet50 for recognition of activity. Eventually, with the combination of these two algorithms our system can identify the posture in one second approximately. Labelled images are passed to CNN for the training and multiple matrix filters are applied for feature extraction, e.g. edges, color and shape. In case of large image size, it is passed to the pooling layer, used for reducing the image size as well as quality eventually.

4.3 Fall Detection

As already explained earlier that, four poses aimed to identify i.e., sitting, standing, lying, and resting/sleeping. For first three poses, identification is easy, and model has been trained directly. However, the lying and resting poses are very similar, therefore, to distinguish between the lying, resting and fallen over, the detection of bed was also included. In lieu of that, a person will be considered as lying/resting, if human is detected on the bed.

4.4 Reporting of Results

In assistive healthcare, mobile applications have been used to facilitate the user [31]. Therefore, our system has an application, which can display the useful information and provide control option for the camera. The information contains statistical data and logs of the posture along with time stamps. The statistical data shows the posture distribution over the time. The application is developed using JavaScript framework known as 'React Native'. This framework allows efficient development with native support in minimal time.

To use the application, user needs authentication before accessing the private information of the user. The login portal of the application is shown in Fig. 3 (a). To access the information on the application, user need to register by providing the personal details i.e., Email ID and name. These details are being used to make the profile of the user and after providing authorization to the user, he/she can access the application and data on the cloud. This application is cable to handle multiple cameras and multiple users. For instance, if there are multiple cameras, only the relevant camera access will be granted to the client or guardian. Moreover, relevant access can be granted to healthcare professional to further analyse the data.

As the user logged in the application, a bar graph is shown on the screen, demonstrating the percentage of posture over the time as shown Fig. 3 (b). Moreover, the information about the recorded posture along with the time stamp is also provided as shown in Fig. 3 (c). Furthermore, user is provided with real time images of the posture recognized and an option to save the required image for later use if permission is granted to that specific user. Considering the elderlies, a warning system has also been introduced in the application. For instance, if the elderly person is sitting in a specific pose for long, which can affect his health, a warning indication will be generated on the application for the caretaker or health professional. Similarly, as per the recommendation of caretaker or health professional, multiple warning can be provided in the application. In continuity of that, there are some health issues in elderly people, in which they keep on repeating some tasks or forget about what they are doing, so, this application warning system can also be helpful for their caretaker or health professional to keep an eye on the user autonomously.

4.5 Methodology for Software Development

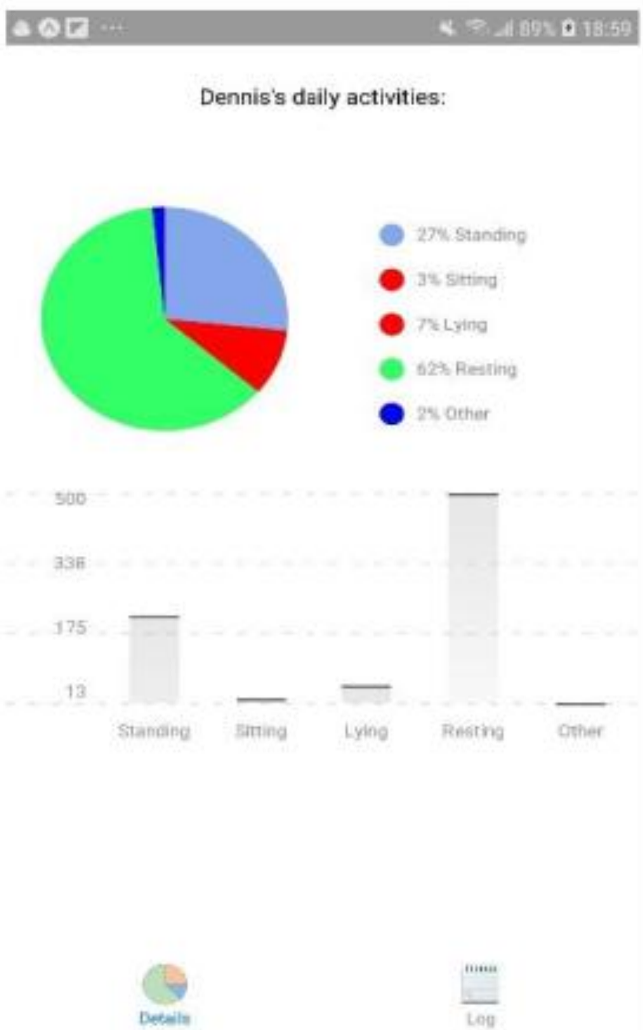
Choosing a software platform is very much dependent on the user application, therefore, initially we consider the usage and user of our application, followed by exploring the critical and essential features of the application.

Considering the requirements of the user, different methodologies were shortlisted and finally a software development model called as 'incremental build model' is adopted. This model is very much similar of 'Waterfall model' except the validation process. The Waterfall model repeat every step multiple time and validates its success, whereas, incremental build model, validates the system for any design change after a full cycle as shown in Fig. 4 [32]. This is also the main logic behind choosing this model. To find the accuracy of the system, our system requires frequent checking of machine learning algorithms. Moreover, we also require user feedback for each build; therefore, the said model provides the effortless debugging and testing with minimal cost.

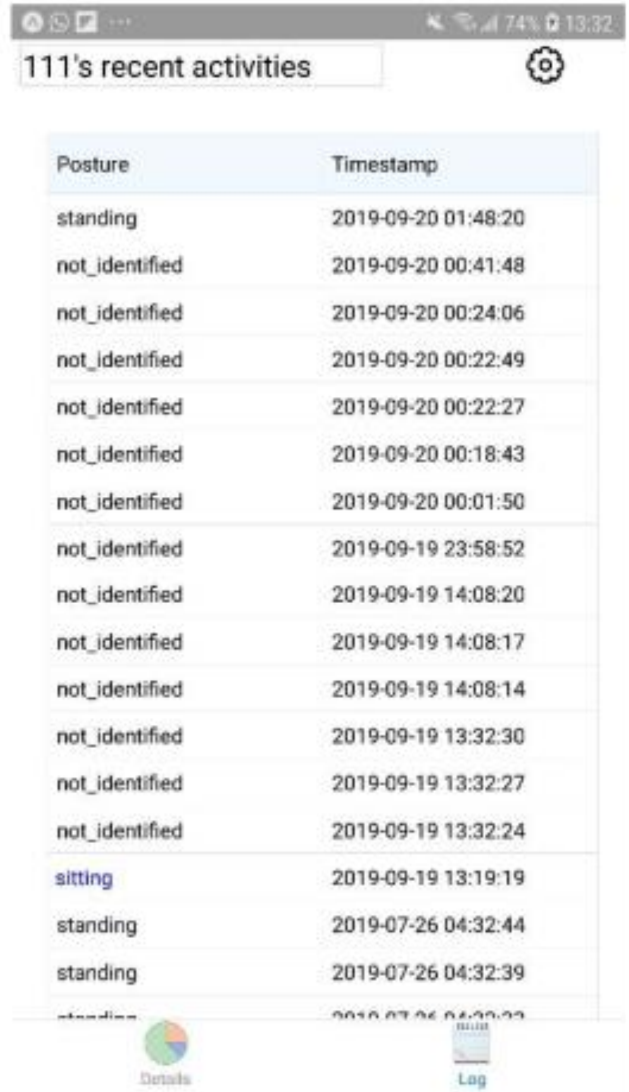
5. Results and Discussion

To test the designed prototype, a small size room has been used, the room has one person living without external interference as shown in Fig. 5. The designed system has not been tested intentionally on elderly people, because of their health concerns. Our main focus was to find out a solution with high accuracy in minimal processing time and also to utilize minimal memory. In lieu of that, our system shows promising results as explained below.

(a)



(b)



(c)

Fig. 3. User interface of smartphone application, (a) login interface, (b) Interface for statistical analysis of user activities (c) Log of posture and activity recognition

Table 1

Images used for training and test dataset

Category	Training	Test	Total
Unoccupied (U)	200	20	220
Face Up (FU)	200	20	220
Face Down (FD)	200	20	220
Left Lateral (LL)	200	20	220
Right Lateral (RL)	200	20	220
Edge (E)	200	20	220

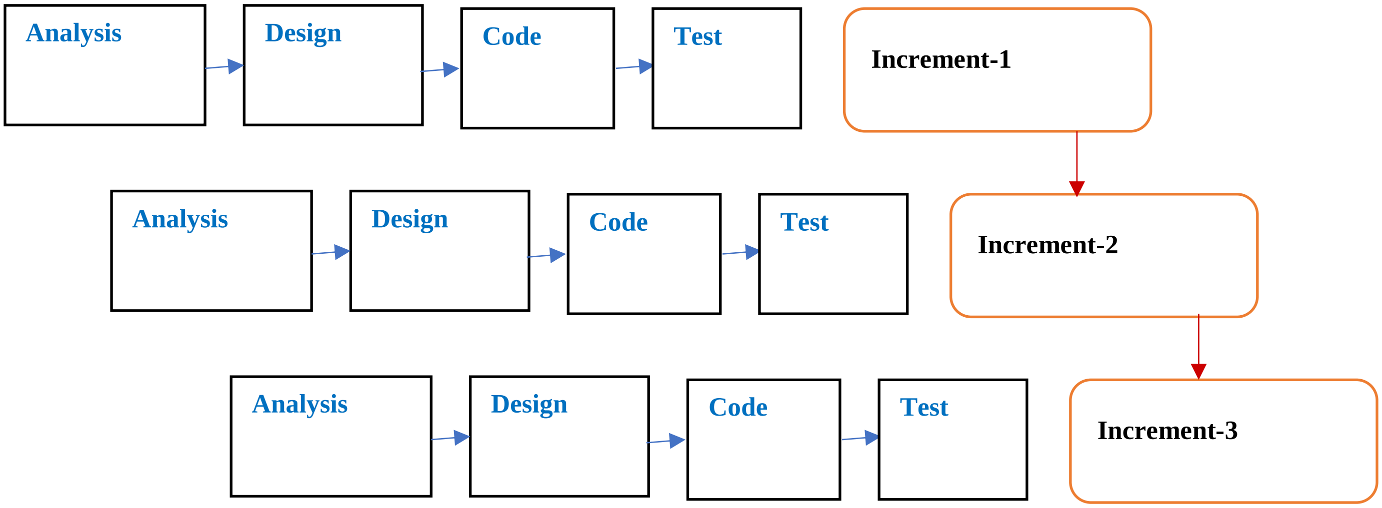


Fig. 4. Incremental build model [27]

(a) (b)

Fig. 5. Floor plan of the environment used for testing, (a) A 3D rendering of the environment, (b) The layout design of the environment

For the training of the dataset multiple images for each posture has been used. To increase the accuracy, the dataset has been trained multiple times and after each training session, the accuracy has been improved, however, it has been observed that by using the more training dataset the system over-tuned. Therefore, for optimal accuracy (above 85%), 200 images for each posture have been used as given in Table 1. The dataset has been divided into training and test data for each posture with no overlapping.

For testing of the system, recordings have been done for each action. To fine out the percentage accuracy of the system, correctly recognized posture by the system is divided with the total number of frames analyzed by the system. The details of accuracy against each posture are shown in Table 2, whereas pictorial examples of each recognized posture are shown in Fig. 6.

Table 2

Percentage accuracy of recognized postures

Posture	Standing	Sitting	Lying	Resting (Sheets)
%Accuracy	90.68%	88.37%	28.42%	71.19%

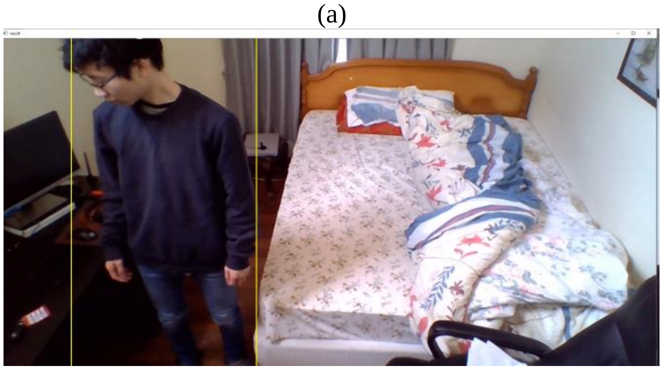
Table 3

Technical specifications of different platforms for detection and recognition.

Platform	Desktop (CPU)	Desktop (GPU)	Raspberry Pi 3 B+
Processing Unit	Intel Core i5-4460 CPU @ 3.20GHz Quad-Core	NVIDIA GeForce GTX 960, Intel HD Graphics 4600	1.4 GHz 64/32-bit quad-core ARM Cortex-A53
Processing Time (s)	~4	~1	>10

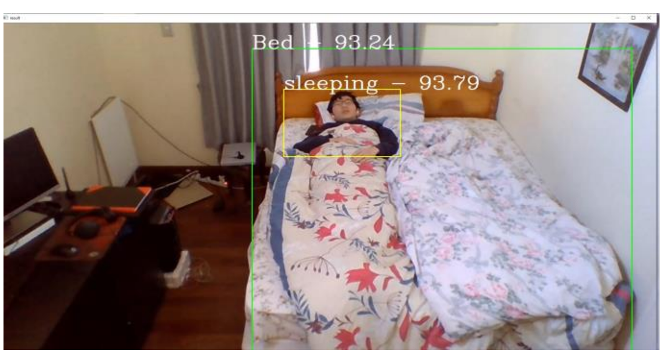
While choosing the platform for the testing of the proposed system, three choices has been considered, (1) a desktop computing system with dedicated Graphics Processing Unit (GPU), (2) a desktop computing system, and (3) a Raspberry Pi based system. A comparison of time taken for posture recognition is shown in Table 3 which was main consideration when choosing the system. As shown in the table, the desktop computing system with dedicated GPU is most time efficient, whereas the Raspberry Pi 3 B+ is least time efficient.

(a)



(b)

(c)



(d)

Fig. 6. Example of the detected and recognized postures with percentage confidence, (a) Example of sitting results, (b) Example of standing results, (c) Example of incorrect lying results, (d) Example of resting results with bed detection

The results shown in Table 2 are for standing, sitting, resting and lying postures and the average accuracy of the system for all the postures is approximately 86%. The accuracy for standing, sitting and resting is quite good, however, the accuracy for lying is concerning. There are different factors influencing the accuracy of designed model. The most prominent factor is light condition in the room. For instance, if the lighting conditions in the room are poor and user is wearing darker clothes, than detection of human and recognition of the posture become difficult. Other factors affecting are occlusion and obscurity. For example, if the entire body of user is not in the view of camera, then the chances of false detection increase rapidly. Moreover, the angle of vision has also an impact on detection and recognition. Specially, in case of lying this factor plays an important role. Lastly, similarity of postures also makes it difficult to recognize the posture accurately. For instance, when the image is cropped after human detection and passed to recognition sever, the similarity between standing and lying cropped image is quite high. Therefore, accuracy of lying posture is low due to this factor. Also, the fall detection works well but there are certain angles that do not suit a standard room environment.

For training the system using ImageAI, a desktop computing system with dedicated Graphics Processing Unit takes over 12 hours to train model for 200 images. Moreover, with a small iteration, whole training process needs to be restarted. Therefore, the best available resources are being used for designing and implementing the system.

As explained earlier, royalty-free images are being used in the training and testing dataset. Therefore, many of those images are taken at a particular angle for specific application. For instance, in case of lying, mostly images are taken while lying flat at ground level. These images are taken in ideal condition, while the real time lying images are bit different, which also contributed to low accuracy for lying and this can be improved in extended research by using better dataset for training.

Similarly, in case of human detection, there are chances of false detection due to many conditions, few of them have already been discussed i.e., lighting conditions, occlusion, obscurity and angle. Therefore, if false human detection occurs, it will eventually contribute to false posture recognition. The dataset (COCO) we used in our system has large images for humans' detection, still due to certain angles, the system does false detection. The most distinctive feature for human detection is face, so, while testing it has been observed that if the user is not facing towards camera the percentage of false human detection increases significantly. The said issues can be seen in Fig. 7, in which the face of user is at an angle, which contributed towards false human detection and eventually false posture is recognized.

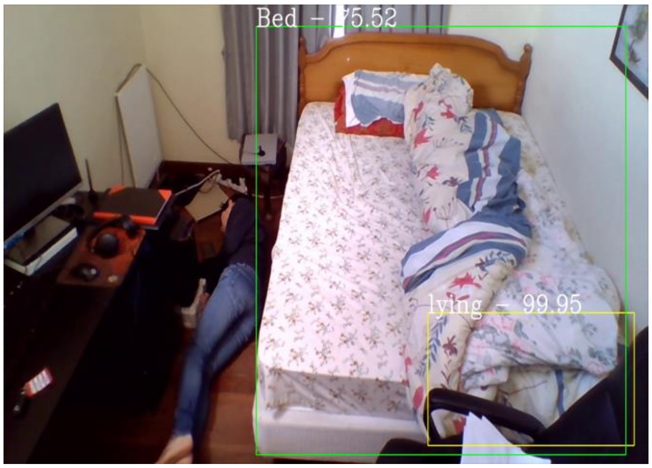


Fig. 7. Example of a failed detection

6. Conclusion

The elderly people face difficulties while living alone is because of their lacking physical and cognitive abilities. Therefore, some time they need aid by others in their daily routine. The proposed system provides a compute-vision based solution as an alternative to not so formidable solution of 24/7 human aid to elderlies. This system not only facilitates the elderlies but also to the caretaker or healthcare worker by providing an autonomous activity supervision and fall detection. The proposed solution is equipped with an application for controlling and acting as a monitoring device. It is also capable to generate alerts in case of fall or other forbidden action. The experimental results of standing, sitting and resting are promising, however, lying results are bit compromised and need further research before implementing in the real world. However, the better results can be achieved with improved dataset and usage of multiple cameras. Moreover, this system is not limited to human activity detection only, it can be improved further by incorporating the equipment, room and sign detections, under the guidance of caretakers and healthcare professionals already taking care of elderly people.

7. Reference

[1] J. Cornwall and J. Davey, "Impact of population ageing in new zealand on the demand for health and disability support services, and workforce implications", Victoria University of Wellington, Ministry of Health, Wellington, 2004.

[2] G. Williams, J. Cylus, S. Barber and T. Roubal, "How will population ageing affect health expenditure trends in New Zealand and what are the implications if people age in good health?", European Observatory on Health Systems and Policies, Economics of Healthy and Active Ageing, WHO Regional Office for Europe, pp. 1-16, 30 November 2020.

[3] M. Parsons, P. Rouse, L. Sajtos, J. Harrison, J. Parsons, and L. Gestro, “Developing and utilizing a new funding model for home-care services in New Zealand”, Health and Social Care in the Community, vol. 26, no. 3, pp. 345-355, 2018.

[4] Statistics New Zealand, "Census QuickStats about people aged 65 and over", Govt of New Zealand, https://www.stats.govt.nz/topics/older-people [Accessed on 13 June 2020].

[5] M. Horne, D. A. Skelton, S. Speed, and C. Todd, “Falls prevention and the value of exercise: salient beliefs among South Asian and white British older adults”, Clinical Nursing Research, vol. 23, no. 1, pp. 94-110, 2014.

[6] S. N. Robinovitch, F. Feldman, Y. Yang, R. Schonnop, P. M. Leung, T. Sarraf, J. Sims-Gould, and Loughin, “Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study”, Lancet, vol. 381, no. 9860, pp. 47-54, 2013.

[7] M. A. Asari and U. U. Sheikh, “Vision based assistive technology for people with dementia performing activities of daily living (ADLs): an overview”, Fourth International Conference on Digital Image Processing, SPIE Kuala Lumpur Malaysia, 2012.

[8] S. Bersch, C. Chislett, D. Azzi, R. Khusainov, and J. Briggs, “Activity detection using frequency analysis and off-the-shelf devices: Fall detection from accelerometer data”, Proceedings of the IEEE 5th International ICST Conference on Pervasive Computing Technologies for Healthcare, Dublin Ireland, 2011.

[9] H. W. Guo, Y. T. Hsieh, Y. S. Huang, J. C. Chien, K. Haraikawa, and J. S. Shieh, “A threshold-based algorithm of fall detection using a wearable device with tri-axial accelerometer and gyroscope”, IEEE International Conference on Intelligent Informatics and Biomedical Sciences, Okinawa Japan, 2015.

[10] A. Buke, F. Gaoli, W. Yongcai, S. Lei, and Y. Zhiqi, “Healthcare algorithms by wearable inertial sensors: a survey”, China Communications, vol. 12, no. 4, pp. 1–12, 2015.

[11] S.-J. Jung, T.-H. Kwon, and W.-Y. Chung, “A new approach to design ambient sensor network for real time healthcare monitoring system”, ”, Proceedings of IEEE Sensors Conference, Christchurch New Zealand, pp. 576-580, 2009.

[12] F. Hu, H. Tang, A. Tsema, and Z. Zhu, “Computer vision for sight: computer vision techniques to assist visually impaired people to navigate in an indoor environment”, in Computer Vision for Assistive Healthcare, Academic Press, pp. 1-49, 2018.

[13] X. Chang, C. Peng, T. Hao, G. Xing and G. Zhou, "iSleep: a smartphone system for unobtrusive sleep quality monitoring", ACM Transactions on Sensor Networks, vol. 16, no. 3, pp. 1-32, 2020.

[14] L. Jeon, and J. Finkelstein. "Consumer sleep tracking devices: a critical review", Studies in Health Technology and Informatics, vol. 210, pp. 458-460, 2015.

[15] W. J. Yi and J. Saniie, “Design flow of a wearable system for body posture assessment and fall detection with android smartphone”, IEEE International Technology Management Conference, Chicago IL USA, 2014.

[16] K. Tang, A. Kumar, M. Nadeem and I. Maaz, "CNN-based smart sleep posture recognition system", IoT, vol. 2, no. 1, pp. 119-139, 2021.

[17] S. Muppavram, N. Patel and M. Nadeem, "Posture alert", in IEEE Region Ten Symposium (Tensymp), Sydney NSW Australia, pp. 213-218, 2018.

[18] D. Mulfari, “A tensorflow-based assistive technology system for users with visual impairments”, Proceedings of the Internet of Accessible Things on - W4A, Association for Computing Machinery, New York USA, pp. 1-2 2018.

[19] B. Kaushik, M. Kumar, C. Bhatanagar, and A. S. Jalal, “A context based tracking for similar and deformable objects”, International Journal of Computer Vision and Image Processing, vol. 8, no. 4, pp. 1–15, 2018.

[20] F. Cardile, G. Iannizzotto, and F. L. Rosa, “A vision-based system for elderly patients monitoring”, IEEE 3rd International Conference on Human System Interaction, Rzeszow Poland, 2010.

[21] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, J. Chambers, “A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment”, IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 6, pp. 1274-86, 2012.

[22] M. Martin, P. L. Chiou, T.-F. Cameron, J. Bas and B. Kevin, "Sleep monitor: A tool for monitoring and categorical scoring of lying position using 3D camera data", SoftwareX, vol. 7, pp. 341-346, 2018.

[23] S. Bhatia, L. Sigal, M. Isard and M. J. Black, "3D human limb detection using space carving and multi-view eigen models", IEEE Conference on Computer Vision and Pattern Recognition Workshop, Washington DC USA, pp. 17-17, 2004.

[24] T. Grimm, M. Martinez, A. Benz and R, “Sleep position classification from a depth camera using Bed Aligned Maps,”, IEEE 23rd International Conference on Pattern Recognition, Cancun Mexico, pp. 319–324, 2026.

[25] C. Torres, V. Fragoso, S. D. Hammond, J. C. Fried and B. S. Manjunath, "Eye-CU: Sleep pose classification for healthcare using multimodal multiview data", IEEE Winter Conference on Applications of Computer Vision, Lake Placid NY USA, pp. 1-9, 2016.

[26] J. Lee, M. Hong, and S. Ryu, Lee, Jaehoon, et al., “Sleep monitoring system using kinect sensor”, International Journal of Distributed Sensor Networks, vol. 11, pp. 1-9, 2015.

[27] W. Huang, A. A. P. Wai, S. F. Foo, J. Biswas, C. Hsia and K. Liou, "Multimodal sleeping posture classification", IEEE 20th International Conference on Pattern Recognition, Istanbul Turkey, pp. 4336-4339, 2010.

[28] M. Martinez, B. Schauerte and R. Stiefel-hagen, "BAM! depth-based body analysis in critical care", Computer Analysis of Images and Patterns, vol. 8047, pp. 465-472, 2013.

[29] M. Martinez, L. Rybok, R. Stiefelhagen, "Action recognition in bed using bams for assisted living and elderly care", IEEE International Conference on Machine Vision Applications, Tokyo Japan, pp. 329-332, 2015.

[30] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. DollárC and L. Zitnick, "Microsoft COCO: common objects in context", 13th European Conference on Computer Vision, Zurich Switzerland, 2014.

[31] A. H. Alamoodi, S. Garfan, B. B. Zaidan, et al., “A systematic review into the assessment of medical apps: motivations, challenges, recommendations and methodological aspect”, Health and Technology, vol. 10, pp. 1045-1061, 2020.

[32] M. Martin, "Incremental model in SDLC: use, advantage & disadvantage", https://www.guru99.com/what-is-incremental-model-in-sdlc-advantages-disadvantages.html [Accessed on 13 June 2020].

