

Article

# Deep-cARe: Projection-Based Home Care Augmented Reality System with Deep Learning for Elderly

Yoon Jung Park <sup>1</sup>, Hyocheol Ro <sup>1</sup> , Nam Kyu Lee <sup>2</sup>  and Tack-Don Han <sup>1,\*</sup>

<sup>1</sup> Media System Lab, Yonsei University, Seodaemun-gu, Seoul 03722, Korea; yoonjung.park@mssl.yonsei.ac.kr (Y.J.P.); hyocheol.ro@mssl.yonsei.ac.kr (H.R.)

<sup>2</sup> College of Computer Science & Software, Chung-Ang University, Dongjak-gu, Seoul 06911, Korea; nkleee27@cau.ac.kr

\* Correspondence: hantack@mssl.yonsei.ac.kr

Received: 27 June 2019; Accepted: 7 September 2019; Published: 17 September 2019



**Abstract:** Developing innovative and pervasive smart technologies that provide medical support and improve the welfare of the elderly has become increasingly important as populations age. Elderly people frequently experience incidents of discomfort in their daily lives, including the deterioration of cognitive and memory abilities. To provide auxiliary functions and ensure the safety of the elderly in daily living situations, we propose a projection-based augmented reality (PAR) system equipped with a deep-learning module. In this study, we propose three-dimensional space reconstruction of a pervasive PAR space for the elderly. In addition, we propose the application of a deep-learning module to lay the foundation for contextual awareness. Performance experiments were conducted for grafting the deep-learning framework (pose estimation, face recognition, and object detection) onto the PAR technology through the proposed hardware for verification of execution possibility, real-time execution, and applicability. The precision of the face pose is particularly high by pose estimation; it is used to determine an abnormal user state. For face recognition results of whole class, the average detection rate (DR) was 74.84% and the precision was 78.72%. However, for face occlusions, the average DR was 46.83%. It was confirmed that the face recognition can be performed properly if the face occlusion situation is not frequent. By object detection experiment results, the DR increased as the distance from the system decreased for a small object. For a large object, the miss rate increased when the distance between the object and the system decreased. Scenarios for supporting the elderly, who experience degradation in movement and cognitive functions, were designed and realized, constructed using the proposed platform. In addition, several user interfaces (UI) were implemented according to the scenarios regardless of distance between users and the proposed system. In this study, we developed a bidirectional PAR system that provides the relevant information by understanding the user environment and action intentions instead of a unidirectional PAR system for simple information provision. We present a discussion of the possibility of care systems for the elderly through the fusion of PAR and deep-learning frameworks.

**Keywords:** Projection Augmented Reality; elderly; home care; assistance; safety; deep learning; pose estimation; face recognition; object detection

## 1. Introduction

Human lifespan has increased owing to recent medical developments. It is expected that approximately 20% of the global population will be 60 years of age or older by 2050 [1]. Furthermore, the number of elderly single-person/couple households is sharply increasing with changing family structures, such as nuclear families, owing to urbanization in modern society. In the past, the elderly

were usually attended to by their families; however, they are currently forced to lead independent lives [2]. There has been increasing interest in social welfare and medical support owing to the increasing aging rate. The social government is increasing social welfare and medical benefits to support elderly self-sustenance; however, most elderly people attend to various problems in their daily life on their own. Moreover, various diseases as well as physical and mental deterioration can occur with aging [3]. Muscle weakness is a representative example of physical deterioration. Muscle strength of people in their 60's decreases to approximately one half of that in their 20s, resulting in significant discomfort in everyday life. In the case of presbyopia, which is another example, the eyes of 60-year-old people can only accept approximately one third of the light that they could in their 20s, accordingly bright places are preferred owing to the reduced pupil diameter. Moreover, presbyopia commonly results in focus loss of nearby objects. Consequently, elderly people experience difficulty in reading information presented in small text or on smartphones. Furthermore, sensory functions begin to deteriorate as people approach their 50s. As the ability to regulate and control body temperature declines with age, the elderly become vulnerable and sensitive to low and high external temperatures [3]. Despite the discomfort in everyday life, the number of elderly people who choose to live at home rather than in dedicated facilities is increasing [4]. However, safety-related accidents occur most often at home. In particular, 25.1% of the elderly experienced falls during the one year. Elderly people living alone experience considerable anxiety associated with falling accidents [2].

Information and communication technologies can be used to assist in the prevention of accidents and to mitigate the social, physical, and mental problems that the elderly generally experience. Several health management systems and elderly support functions are being developed by fusing various sensor and network technologies. An example is smart homes for the elderly and the technologies, tools, and systems associated with active assisted living [5]. In these homes, vital signals, behavioral information, and other contextual information can be collected through motion, radar, pressure, and vision sensors. The data collected can be analyzed to monitor the health and behavior of the elderly. In particular, vision and video sensors are most commonly used for elderly care. Vision cameras are installed inside the walls and ceilings of a house to monitor the locations and behaviors of the residents [6]. Several studies on the prevention and recognition of falls, including rehabilitation training after fall accidents, have been conducted using joint information received through RGB-depth cameras [7].

This study proposes healthcare support functions for the elderly in situations that can cause discomfort in daily life. Specifically, we employ an augmented reality (AR) system and deep-learning technology based on vision cameras and projection. This paper is organized as follows. Section 2 presents related work on support systems for the elderly using AR and deep-learning technologies, which are classified based on the display type. Moreover, trends in elderly care systems are analyzed, and the distinctiveness of the proposed system is described. Section 3.1 introduces the hardware components used in the proposed system. In addition, technologies for a projection-based AR (PAR) environment in Section 3.1.3, including deep-learning technologies for elderly support in Section 3.1.4, are described. Experiments are also presented in Section 4 to evaluate the integration of deep-learning technologies. Section 5 describes the scenarios and the interaction design for implementing a PAR application for the elderly. Finally, Section 6.1 concludes the paper.

## 2. Related Work

Recently, there is increasing interest in developing support systems that slow aging or continuously monitor the health conditions of the elderly rather than developing treatment methods. Studies on the development of systems that monitor the current condition of an elderly person include a study concerned with the determination of falling using vision information and deep-learning technology [8] and a study on behavior recognition in daily life [9]. Furthermore, a recent study proposed a system that monitors the current emotional state of an elderly person through emotion recognition. The system then conveys the emotional state information to the caregiver or hospital and

accordingly provides treatment and mental assistance [10]. Furthermore, a guideline for diet therapy by recognizing foods and drinks using deep learning was proposed in [11].

In addition to monitoring systems, AR systems that allow the elderly to access a wide range of media information and content are currently being developed so that physical and mental deterioration may be mitigated. The display type of AR systems may be either a head mounted display (HMD), a mobile device providing virtual information, or a projector (in PAR).

Research on therapeutic systems that employ HMD AR for patients with dementia and Alzheimer's disease is underway. A study that focuses on elderly with dementia is currently underway. In [12], dementia patients are assisted in their daily activities by virtual information so that their physical and cognitive functions may improve by repeated use. Furthermore, an educational system was developed to assist those with Alzheimer's disease by improving their short-term and spatial memory and mitigating mental decline [13]. A HMD AR game was developed for Alzheimer's diagnosis [14]. In addition, a system for providing advice information on medicine through HMD was developed [15]. Few studies have been concerned with elderly care systems that use an HMD AR system and deep-learning technology. In the case of HMD, the range of information delivery is limited because the field of view is approximately 30° (Microsoft, HoloLens). An HMD system can also be inconvenient because the user should continuously wear it for consistently receiving information. HoloLens is fast and accurate spatial mapping, via an IR depth camera and visible light cameras, and dual HD translucent display [16]. Although HoloLens supports visual and tactile interactions with virtual objects, this interaction method can be difficult for the elderly to use [13].

A mobile AR system is relatively easy to develop and can provide a friendly user interface (UI) because it uses mobile devices. For example, applications have been developed that employ a mobile device's camera and provide related information through the device's display when the user uses a medicine case [17,18]. Developed for mobile AR games for health promotion through the user-friendly mobile devices [19]. Furthermore, there exists a mobile AR system that recommends the best location for installing safety bars to prevent accidents [20]. However, mobile AR systems can increase fatigue because they are hand-held, and the user should continuously hold a smart device. Furthermore, these devices are less realistic than AR headsets because virtual objects and information are received through mobile device displays.

A PAR system can mitigate the problems of existing studies and deliver information to a large number of users as well as ensure high realism by projecting virtual information directly onto the real world; thus, the system can naturally assist users. In [21], a PAR table-top system that provides daily care to patients with memory problems was studied. In [22], a system was developed that employs image and text recognition to control a smart home using PAR. A home-care system that enables the elderly to exercise and enhance their physical abilities through projections was investigated in [23]. Most PAR systems are either fixed mount or table-top [21,22]. The UI uses image recognition and projection, but several problems have been reported in the application of these systems.

Therefore, independent studies are underway to develop elderly care systems based on deep learning and PAR technologies shown in Table 1. However, there is limited research on home-care systems through the fusion of a PAR system that relays virtual information using deep-learning technology so that the system can be used in various environments.

This study proposes such an elderly care system that dynamically provides information using projections. This information can aid elderly people in their daily lives. By combining deep-learning object-recognition technology with PAR, the system can be immediately used in predefined locations that are predefined or in a laboratory environment. Furthermore, deep-learning pose estimation (rather than simple monitoring) was applied for a more detailed and accurate determination of the elderly's condition. This can prevent accidents by detecting abnormal conditions, such as falling and tripping. Furthermore, daily-life scenarios can be provided to patients exhibiting memory-related problems through the real-time recognition of typical real-world objects inside a house.

**Table 1.** Related work for elderly using AR system.

Research Authors (Year)	Display	Purpose	Application	UI	Characteristic
Wolf et al. [12] (2018)	HMD (Wearable)	Daily support	Cooking	- Hand gesture - Host PC	- 3D space reconstruction - Spatial mapping
Aruanno et al. [13] (2019)		Mental care	Game	- Hand gesture - Voice - Controller	- 3D space reconstruction - Spatial mapping
Vovk et al. [14] (2019)		Mental care	Game	Hand gesture	- 3D space reconstruction - Spatial mapping
Ingeson et al. [15] (2018)		Memory assistant	Medication management	None	Cloud-based image recognition
Lera et al. [17] (2014)	Mobile device (Hand-held)	Memory assistant	Medication management	Mobile UI (MUI)	Feature-matching-based image recognition
Chae et al. [18] (2016)		Memory assistant	Medication management	MUI	Feature-matching-based image recognition
Schlomann et al. [19] (2019)		Physical care	Game	MUI	Location-based AR
Bianco et al. [20] (2016)		Safety	Augmented safety bar	MUI	Feature matching-based image recognition
Yamamoto et al. [21] (2015)	Projector (Fixed Table-top)	Daily support	Home control	Spatial UI (SUI) with finger ring	Spatial touch detection
Pizzagalli et al. [22] (2018)		Daily support	Home control	SUI	- Cloud-based image recognition - Spatial touch detection
Vogiatzaki et al. [23] (2019)	Projector and Monitor	Mental and Physical care	Game	- Voice - Body gesture - Host PC	- EMG sensing - Body movement recognition
Petsani et al. [24] (2018)		Mental and Physical care	Tele-stretching	- Voice - Body gesture	- Body movement recognition
Our propose Park et al. (2019)	Projector (Portable)	- Safety - Daily support - Memory assistant	- Monitoring - Home control - Medication management	- MUI - SUI	- 3D space reconstruction - Spatial mapping - Deep learning-based Pose estimation, face recognition, and object detection - Spatial touch detection

### 3. Material and Methods

This study proposes an elderly care system by applying deep-learning technology that can build a PAR environment for every space in which users are present (Figure 1). The proposed system aims to dynamically provide appropriate information to the elderly using pan-tilt mechanism PAR hardware (Section 3.1.1), constructed PAR environment (Section 3.1.3), and designed UIs (Section 5.2). As shown in Figure 1—upper part, deep learning-based pose estimation technology was applied to enable detecting abnormal conditions. To support long-term memory, deep learning is applied to the face recognition module. Furthermore, daily-life scenarios can be provided to patients exhibiting memory-related problems through the real-time object detection based on deep-learning framework. The technologies in each deep-learning module are described in detail in Section 3.1.4.

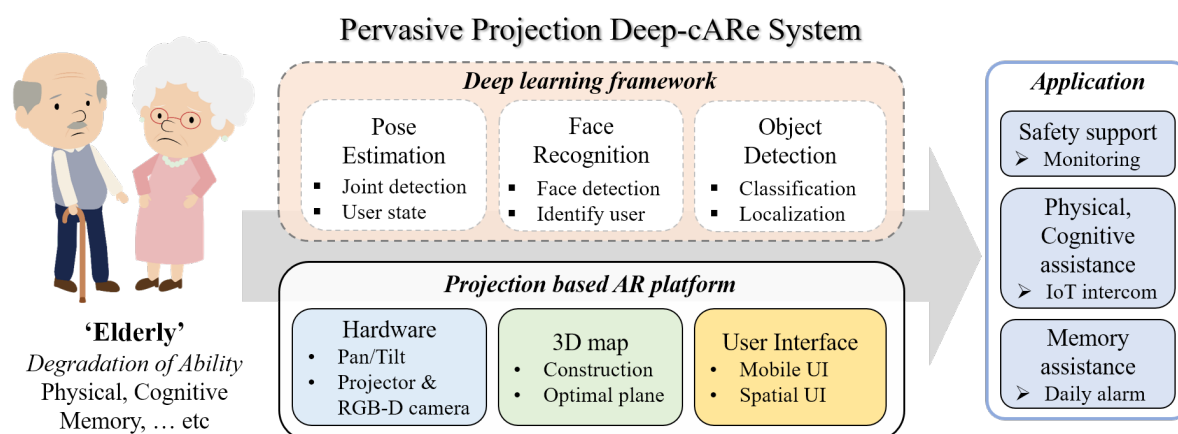


Figure 1. Overview of pervasive projection Deep-cARe system for the elderly.

### 3.1. Deep-Care System Implementation

#### 3.1.1. Hardware Configuration

The proposed system is a projector-camera unit system that can rotate 360°. As shown in Figure 2, a projector (*Sehyun<sup>TM</sup>*, 330 ANSI) that projects information, content, and a UI to users and an RGB-depth-sensing camera (*ASUS<sup>TM</sup>* Xtion pro live) that detects the surrounding environment, including user information, are mounted at the top. The system then constructs a 3D map using information on the surrounding environment, which is obtained from the camera. Furthermore, the system also receives information related to objects around the user, user location, and user actions. It can control two pan-tilt servo motors using Arduino and rotate 360° horizontally and 180° vertically. The Hololens 2 from Microsoft, an HMD device that can provide an AR environment, is priced at \$3500 per device. The HMD application for elderly is developed separately. In general, least 6 to 8 units multiple projector-cam are required to provide a PAR environment to the surrounding use’s space. However, using the pan-tilt mechanism, the system can recognize the entire space with one piece of hardware and provide dynamic information to the user, thereby reducing the cost savings associated with the construction of a PAR environment for user surroundings. The elderly experience a deterioration of light-accepting ability and prefer bright spaces owing to reduced pupil diameter. In addition, they often experience presbyopia and lose the ability to focus on near objects. To complement enhance visual ability, the appropriate information and applications are provided through the projector. As this system is compact (W × H × D, 16 × 22 × 17 cm) and light of weight (1.3 kg), it can be easily carried and placed in various locations.

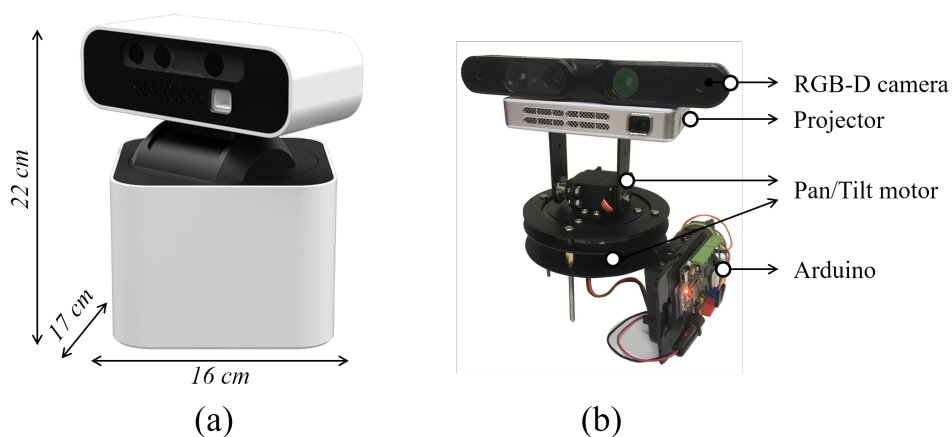


Figure 2. Deep-cARe hardware (a) Housing (b) Components.

### 3.1.2. Software Configuration

Deep-cARe includes input/output components in the form of hardware and a series of processors that extract only the required information from the input data. The overall Deep-cARe system architecture is shown in Figure 3. The PAR module for providing a PAR environment to the user. The deep-learning module for analyzing and recognizing information related to the user in the PAR space. The user interface module for system control. First, the PAR module comprises an RGB-depth and an intercom camera. Furthermore, the PAR module includes plane detection software for detecting the appropriate projection space from actual input space information. The module generates a PAR environment from the space around the user and includes an Arduino module for controlling pan-tilt servo motors that provide information concerning the detected plane. In addition, the module uses a projector as an output component. The deep-learning module performs pose estimation for extracting joint and current state information from the input user data, face recognition, and object detection. The UI module comprises short range (SUI) and middle-long range (MUI) interfaces for intuitive and convenient interactions in various situations.

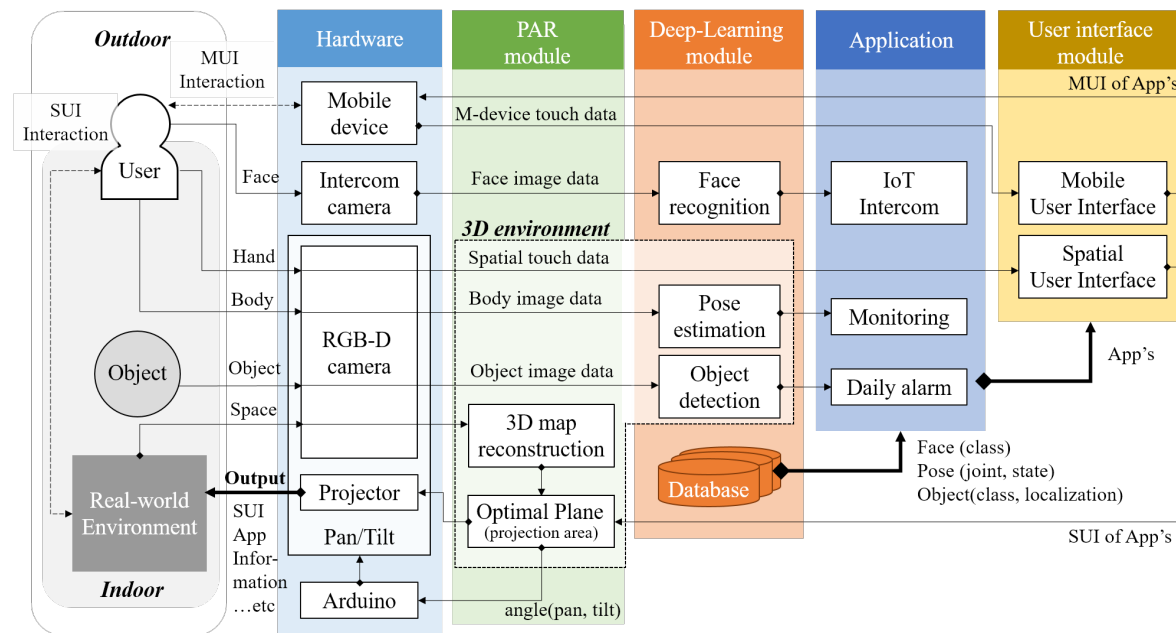


Figure 3. Deep-cARe system software component module flow diagram.

### 3.1.3. PAR Module

There are basic technologies to efficiently provide Deep-cARe system living support applications in the real environment. First, the space information is reconstructed in 3D through color and depth image. The optimal plane that can be easily and conveniently used by the elderly is extracted by analyzing the reconstructed 3D space. The optimal plane extracted through projection is provided for the information and application that are meaningful to the user.

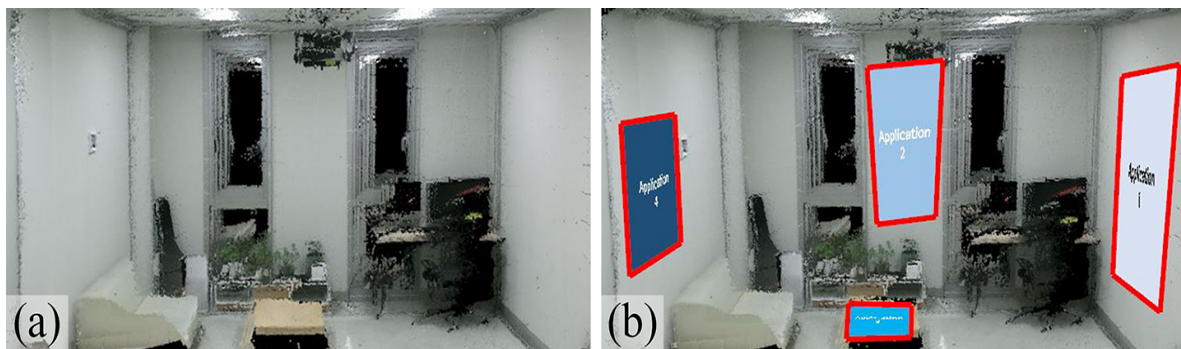
#### (a) 3D space reconstruction

To present dynamic information to the elderly, a 3D space should be constructed from the real space. To receive space information of an actual environment and reconstruct a 3D map, custom-made hardware, and map-construction technology based on a 3D point cloud are applied. A feature-matching method, which extracts features from the input images and calculates the pose by comparing them to the features of the previous frame, is applied. The 360° surrounding space information is obtained from an RGB-depth camera by setting the servo motor rotation of the pan direction to 10° units. Features are detected using the feature detection algorithm through the acquired color and depth input frames. Matching is then performed by creating

feature descriptors, and the FAST [25] and BRISK [26] algorithms are applied to increase computational speed. The final pose of the current frame is obtained by using the RANdom SAmple Consensus (RANSAC) algorithm [27] on the 3D world coordinate point value of the previous depth image and the two-dimensional (2D) matching points of the current color image. The point cloud information of the current frame is then rotated and translated according to 3D world coordinates. 3D space map reconstruction result as shown in Figure 4a.

(b) Optimal plane selection

In this step, the region that can be projected is extracted to recommend the optimal projection region in the reconstructed 3D map to the user. Depth segmentation is first performed using depth information of the reconstructed 3D map. The plane area is detected by executing the randomized Hough transform algorithm [28] based on the segmented areas. Initially, in the hypothesis step, a model that satisfies the sample data is created by randomly selecting three points at certain distances from the detected point. Subsequently, in the verification step, the plane model is evaluated after the inliers are extracted. If the number of inliers is greater than that of the existing plane, the plane of the largest area is detected by updating a new model. The final projection location is selected by minimizing the projection distortion using the normal vector of the maximum area. Meaningful plane information is extracted for projection from the 3D environment point cloud through this plane detection process. Final optimal planes are represented red boxes as shown in Figure 4b.



**Figure 4.** (a) 3D space map reconstruction (b) Application registration in optimal planes (red box).

### 3.1.4. Deep-Learning Module

Realism and immersion can be ensured by providing virtual information to the real space through PAR technology. However, there are several challenges in providing a PAR system to the elderly. There are many difficulties, deep-learning frameworks are applied. This enables recognition of the current user state, identification, and real-world object. Furthermore, external information can be subsequently used.

(a) Pose estimation

To facilitate monitoring system provision and perform projection on the optimal plane, the user location should be recognized in real time. To estimate the location and pose of the user, PoseNet [29] was used, which is an open-source deep-learning model. PoseNet can track joints and uses 2D images as input. The recognized joint information is converted into 2D coordinates, and the z-axis value (depth information) corresponding to the 2D coordinates is obtained through the depth camera. There are two methods for recognizing specific user states such as falling and tripping using the obtained joint information. First, a study on action recognition in 3D video [30] and a study on skeleton extraction [31] implemented a deep-learning model that

can recognize falling. However, to recognize motions, it can be challenging to crop only the target motion and use it as an input. Therefore, although motions can be recognized with high accuracy in a laboratory environment, it can be difficult to apply the model immediately to a real environment.

In this study, the proposed system was constructed using a second method -rule-based method- according to the joint information detected by the PoseNet deep-learning model. The rule-based method can configure the system by defining cases where the user’s joints are in unusual states. Abnormal states can be recognized by the defined states of the joints and are used for urgent situation alerts such as when the head is lower than other body parts and remains in that state or when most joints stay near the floor (Figure 5 shows that rule-based user state). An abnormal state is identified if the location of the user is the living room or kitchen rather than the bedroom where lying down occurs (Figure 5 shows that space recognition in user). After distinguishing an abnormal state, an alert is sent to the family or rescue team. However, the accuracy of the user’s state recognition can be low only with the joint information detected based on the vision information input through the RGB-D depth camera. To prevent misreporting because of erroneous recognition, the accurate state is determined through the feedback UI response of the user after the abnormal state is recognized (Figure 5 shows that user feedback check state)).

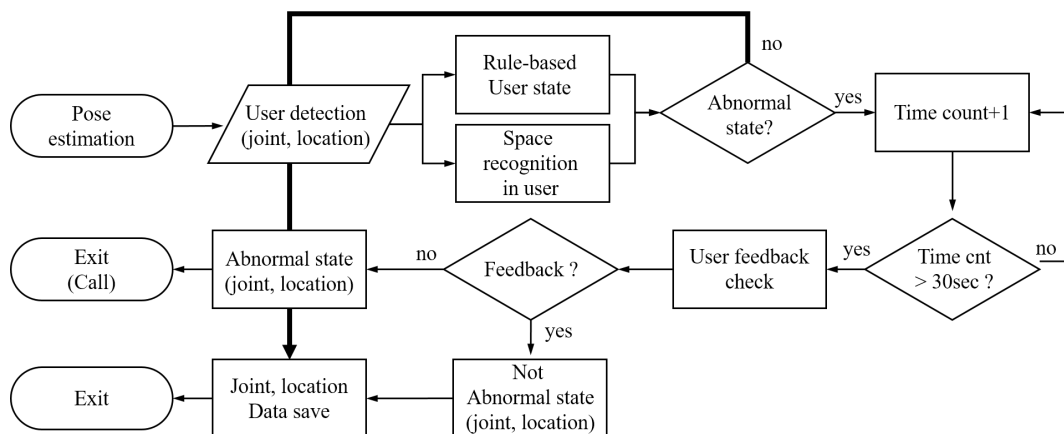


Figure 5. Abnormal state decision flow chart.

(b) Face recognition

Ordinary elderly people do not suffer severe memory decline, as in dementia or Alzheimer’s disease but their ability to identify faces is deteriorated [32]. In particular, they have difficulty identifying people quickly. Erroneous identification of people can result in omission of visiting benefits and delivery services for elderly welfare and may lead to home invasion crimes. To prevent this and support long-term memory, deep learning is applied to the face recognition module. The deep-learning module used was FaceNet [33]. This model can operate at 12 fps near real-time face recognition. An intercom environment was set up to support remote door opening and closing in this scenario. Visitor identification is performed in real time using an image input to the RGB camera attached to the intercom. The images of the faces of likely visitors, such as family, acquaintances, and hospital personnel, are labeled and stored in advance. For previous irregular visitors, such as delivery personnel, the face images are saved as unlabeled data. The face recognition results are largely classified into three types. Family and acquaintances whose data are labeled and stored, are classified as “label name,” people corresponding to unlabeled data are classified as “undefined,” and those who have not



visited and do not exist in the data are classified as “unknown.” This provides information regarding regular, irregular, and new visitors. This will be described in detail in Section 5.1.

(c) Object detection

There are many challenges to use the Deep-cARe system with real-time information in a real (non-predefined) environment. This problem is resolved by applying deep-learning object detection. The proposed system performs object detection for real world objects that may exist inside the house in which the elderly live. To provide information, content, and an appropriate UI for real-world objects in real time, YOLO v3 [34] with a fast processing time (45 fps based on GTX 1080Ti) was used. MS COCO [35] and Open image were used for the dataset. Object detection is performed in real time through the color image from the camera. In this study, a scenario was designed for object detection based on medicines, door, and windows, which are objects that exist in a house and those that can provide the necessary functions to the elderly. This will be described in detail in Section 5.1.

#### 4. Performance Experiment of Deep Learning With Par

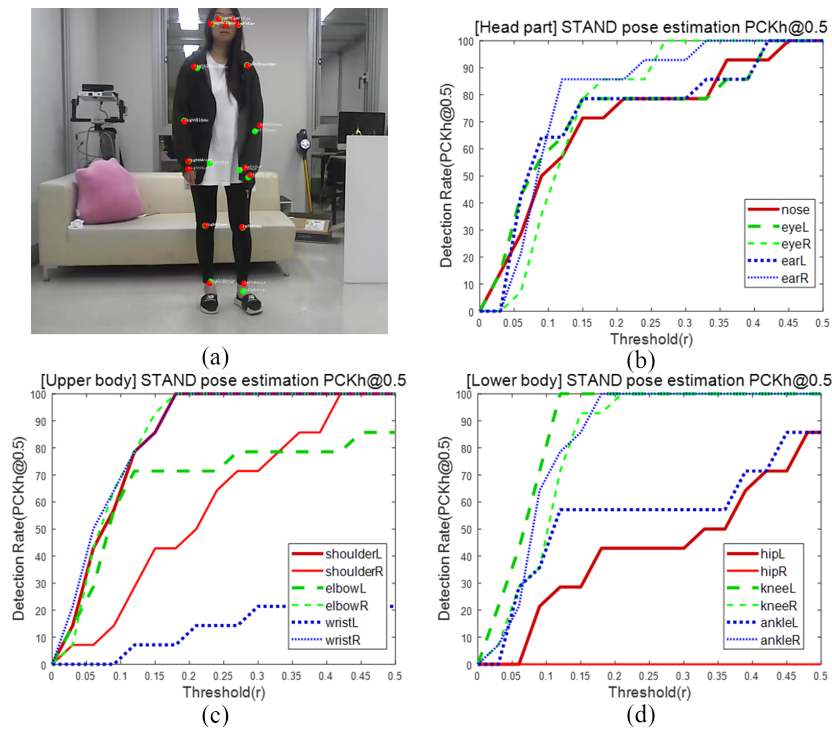
An experiment was conducted to assess the integration deep-learning technology into the PAR system developed using the proposed hardware. To apply the proposed Deep-cARe system to a real house and evaluate its usability and applicability, the experiment was conducted in a laboratory environment in which the interior of a studio apartment was reconstructed. The testbed was in a cubical room of dimensions  $4 \times 3 \times 3$  m (width  $\times$  height  $\times$  depth). It consisted of a living space comprising two windows, a sofa, coffee table, and potted plant. Experiments were conducted to evaluate the performance of the deep-learning object detection in terms of the interaction with internal objects of the house, including pose estimation and face recognition. Experiments on pose estimation and face identification, which require user input, were restricted to a single user. For object detection, the object detection rate (DR) and precision were obtained according to changes in the distance between the object and the camera.

##### 4.1. Pose Estimation Performance

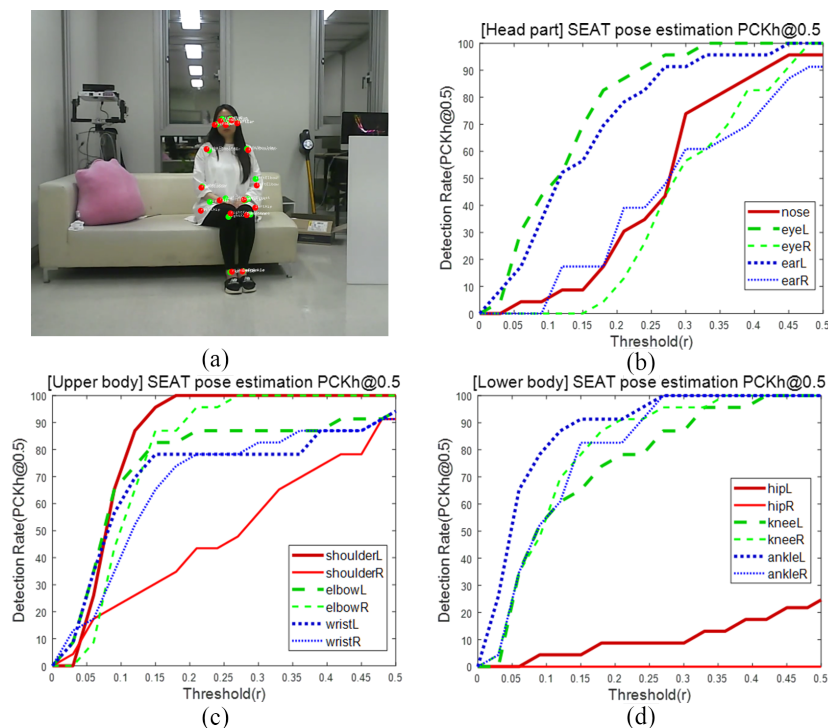
The joint information and user location can be derived through pose estimation. For immediate use of this application in a real-world environment, an experiment was conducted to measure the estimation performance based on the occlusion of the joint information. The input images of a single user were used. Considering four states: “stand,” “seat,” “hide,” and “lie,” 100 frames (25 fps  $\times$  4 s) were used according to each state. The information on 17 joints was estimated. Based on the input images, the ground truth for the joint information was set in advance for each frame. To calculate the precision of the prediction, the percentage of correct keypoints head (PCKh) [36] evaluation metric was used. Based on PCKh@0.5, the standard threshold was set to half of the head diameter (the distance between the left and right ears).

First, in the “stand” state in Figure 6, which was the most normal case, at  $r = 0.5$ , the average precision of all joints was 86.98%. As shown in Figure 6b, the head part exhibited higher precision as the threshold range increased. However, “leftWrist” and “rightHip,” which were occluded by the user’s clothes and other body parts, exhibited low accuracy. Therefore, it was determined that they are highly vulnerable to occlusions.

Subsequently, the “seat” state (Figure 7) demonstrated high precision with an overall precision of 87.47%. In addition, for the “seat” state, the face part exhibited high precision because the front side was directly faced, and the precision of the upper body joint detection was 95.65%, which indicates that the joint information was estimated accurately and stably. However, for the hip part, which was occluded by an arm and contact with the sofa, the average precision was 13.04%. Nevertheless, the lower body excluding the hip had the highest precision 100%.



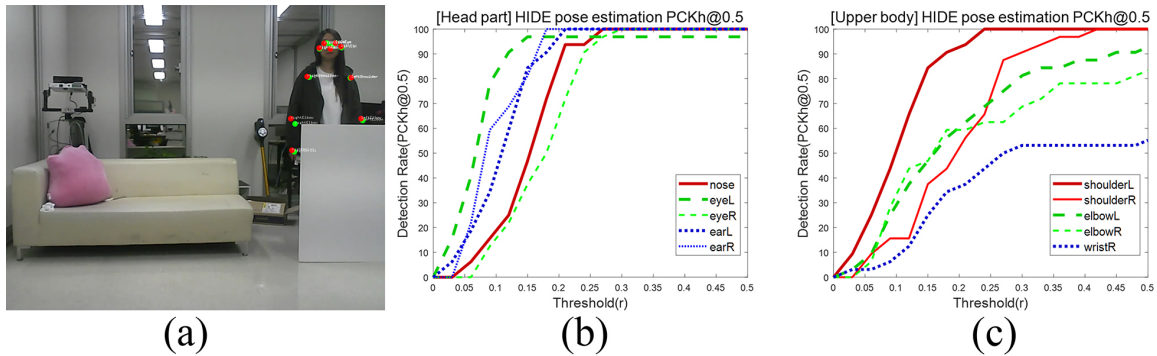
**Figure 6.** “Stand” pose estimation. (a) Ground truth joint (red point) and PoseNet predicted joint (green point). (b) Head part accuracy. (c) Upper body accuracy. (d) Lower body accuracy.



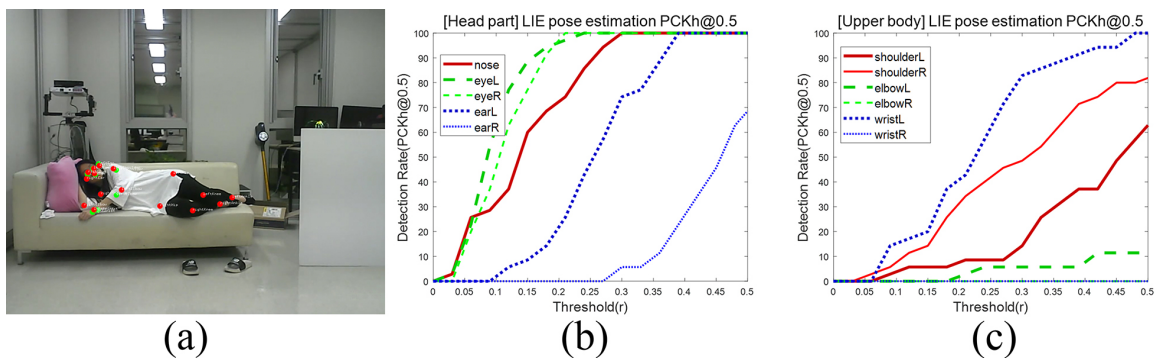
**Figure 7.** “Seat” pose estimation. (a) Ground truth joint (red point) and PoseNet predicted joint (green point). (b) Head part accuracy. (c) Upper body accuracy. (d) Lower body accuracy.

The “hide” state (Figure 8) was applied for performance evaluation in terms of physical occlusions. The average precision of the estimated joints was 93.13%, which is high, although the lower body and left wrist were not estimated at all. Despite the application of deep-learning model, difficulty in joint estimation was experienced because owing to physical occlusions. A “lie” state (Figure 9) was

used for pose estimation performance evaluation considering falls that occur among elderly people. However, the results demonstrated that only the face and upper body were estimated. This was because although the pose estimation itself failed for the test subject, the state of no movement was continuously input. Nevertheless, it appears that the joint information can be estimated with high precision through a series of movements.



**Figure 8.** “Hide” pose estimation. (a) Ground truth joint (red point) and PoseNet predicted joint (green point). (b) Head part accuracy. (c) Upper body accuracy except left wrist.



**Figure 9.** “Lie” pose estimation. (a) Ground truth joint (red point) and PoseNet predicted joint (green point). (b) Head part accuracy. (c) Upper body accuracy.

By using experiments based on the user state, it was confirmed that the joint estimation for the facial part exhibited high precision. Specifically, the facial part joint that is close to the floor surface or an unusual location is used to determine an abnormal state. The overall observation results of pose estimation using PoseNet in a real-world environment exhibited sensitivity to occlusions, confirming that when a user moves quickly, tracking the detected point cannot be performed immediately. However, when all joint information, namely that of the front and rear sides, was accurately input, high precision was observed. Furthermore, because the estimation precision of the face is particularly high, it is used to determine an abnormal user state.

#### 4.2. Face Recognition Performance

Face recognition of a visitor should first be performed to deliver visitor information to an elderly person before the door may be remotely opened. Through a precision experiment involving face recognition, a performance test was conducted to determine whether face recognition can be used in a real environment. First, 30 face images of a user were registered in advance (“label name”). In addition, 15 face images of a user were registered in advance, but no image information was applied (“undefined”). Finally, a user with no face images (“unknown”) was set. One single user was captured by the camera at a time. The face recognition experiment was conducted by changing the front and side inputs and angle of the face and by applying conditions such as wearing of glasses and occlusions

to the face. The experiment was performed on 300 frames (180 frames for changes in the face angle, 60 frames for face occlusions, and 60 frames for the wearing of glasses) per person. As the experiment was conducted to verify whether the deep-learning model for face recognition can immediately reflect a real environment, processes such as biometric system authentication after face recognition were ignored (i.e., true negatives, such as an incorrect recognition and authentication denial, and false negatives, such as a correct recognition and authentication denial, were excluded). In the experiment, the precision measurement was targeted for a single user. Moreover, when an accurate recognition was successful, it was classified as a successful recognition (true positive, TP); however when an incorrect class was recognized, it was classified as an incorrect recognition (false positive, FP), and when the face of the test subject was not recognized, it was classified as a failure (miss). In this manner, the DR was measured. Furthermore, the results were classified by dividing the angle change, face occlusions, and wearing of glasses in all 300 frames. The DR was calculated using the ratio of true positives to the total number of frames, and the precision rate was calculated using the ratio of true positives to the detected frames (sum of the true and false positives).

For a “label name” in advance, Refer to Table 2. Despite changing the face angle between  $-90^\circ$  and  $90^\circ$ , the DR was 87.14% and FP not occurred. With eyeglasses, the DR was 95.22%, confirming that the face recognition function was not significantly affected. However, when a major part, or more than one half of the face was occluded (Figure 10b), the DR was approximately 48.64%, demonstrating a considerably large difference. In the case of “undefined” with information from only 15 images, the precision decreased noticeably. When the angle was changed, a DR of 74.65% and a precision of 83.95% were observed as shown in Figure 10c,d). For face occlusions, the DR was 36.07% and the precision was a mere 9.09%. Finally, in the case of “unknown” users, the average precision was 96.12%, which as high as in the “label name” class. This demonstrated high recognition results when there was no matching information in the stored image data.

Table 2. Face recognition result.

Class	“Label Name”			“Undefined”			“Unknown”		
	Angle	Occlusion	Glasses	Angle	Occlusion	Glasses	Angle	Occlusion	Glasses
TP	157	29	56	113	2	17	136	30	60
FP	0	0	1	22	20	43	1	4	0
Miss	23	31	3	46	38	0	43	27	0

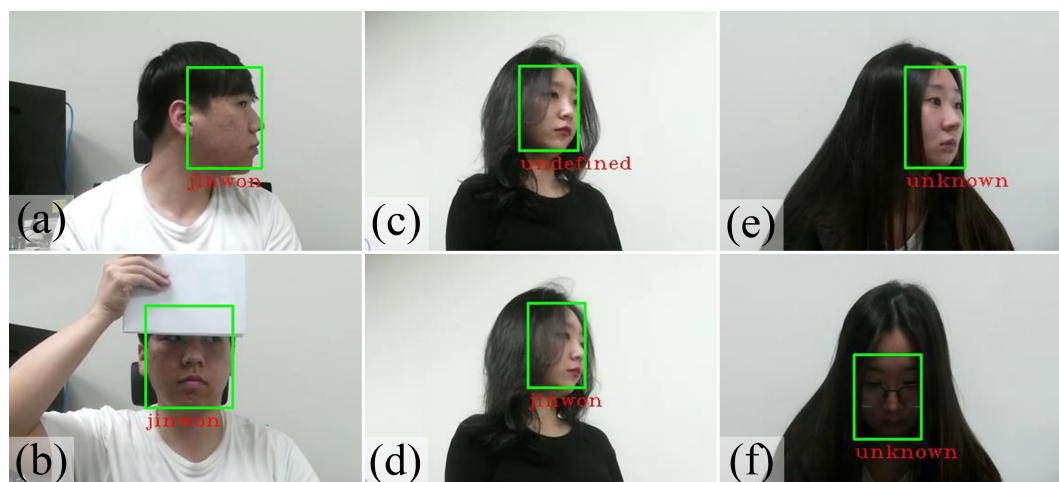


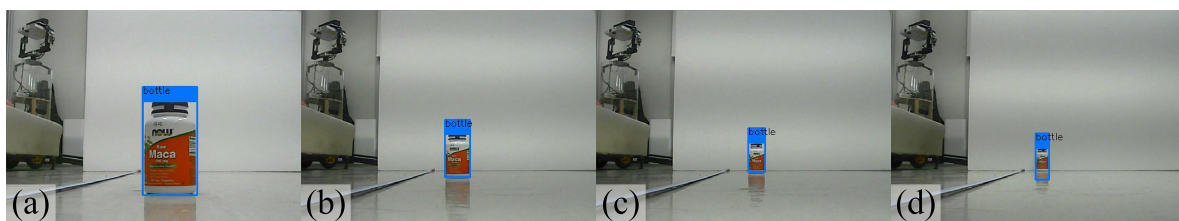
Figure 10. Face recognition result. (a) “label name”,  $90^\circ$  rotation TP. (b) “label name”, face occlusion TP. (c) “undefined”,  $60^\circ$  rotation TP. (d) “undefined”,  $60^\circ$  rotation FP. (e) “unknown”,  $60^\circ$  rotation TP. (f) “unknown”, glasses on TP.

From the experiments, it was confirmed that in the case of unregistered users with 15 face images stored in advance, the precision decreased noticeably. This is because learning was not achieved owing to differences between the information concerning the images stored in advance, and it was confirmed that this can be resolved by increasing the number of stored face images. In general, 100 frames can be acquired from one visit based on the time of visit shown on the intercom. For an unregistered visitor, the number of frames collected for the face image was set to 100. Moreover, as FaceNet is based on color images, it exhibited difficulty in achieving smooth face recognition in a dark environment. However, intercoms are usually installed so that the face of the visitor is clearly visible through an illumination sensor installed in the hallway or entrance. Therefore, the experiment verified that by using a deep-learning face recognition model, the appropriate information can be provided, and the actions of elderly can be easily determined.

#### 4.3. Object Detection Performance

The precision of object detection is based on the environment and the distance from the system, the possibility of an implementation of the scenario can be experimentally determined. To provide notification and information for important scheduling to elderly people in daily life, objects that should be detected are selected. First, elderly patients have difficulty in taking medication at the appropriate time. By performing object detection on a medicine bottle, it is expected that patient compliance can be increased using an AR projection as a medicine-taking notification near the medicine bottle at the appropriate time. Furthermore, through objection detection of windows and doors, status information regarding the interior of the house can be obtained before the subject leaves the house, and guidance can be provided for taking the appropriate action. To verify whether this function can be smoothly performed when the proposed platform is used, a precision experiment on object detection was conducted in a real environment, where the correctly and incorrectly classified cases were categorized as TP and FP, respectively, and when the bounding box was not applied to an object, a detection failure (miss) was recorded. The experiment was conducted by changing the distance between the system and the object, and the detection and precision rates were calculated in terms of the distance in the 375 frames using the same proportional expression as in the face recognition experiment.

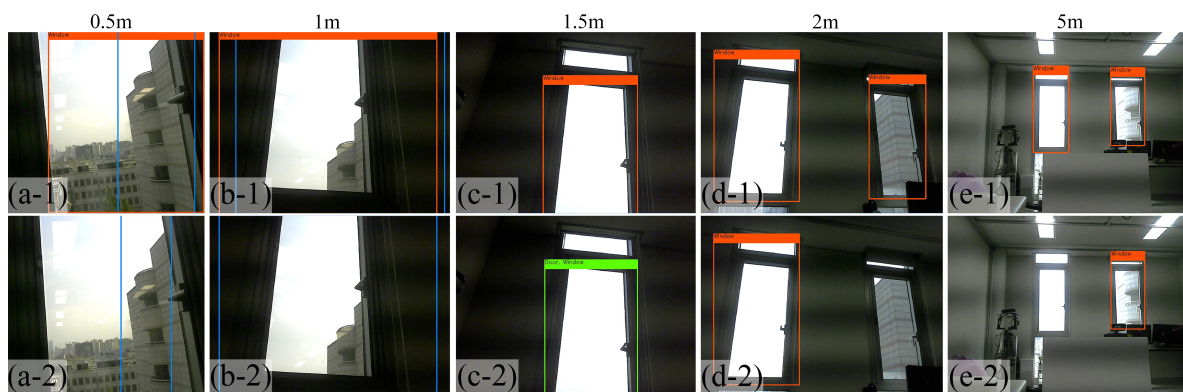
For a medicine bottle (Figure 11), the COCO dataset was used. When the distance from the system was 0.5 m, 100% detection precision was observed, and when the distance was 1 m or larger, the DR was 88.00%. Furthermore, when the distance was 1.5 m, the DR was 86.93%, and when it was 2 m, the DR decreased sharply to 30.13%. When the distance was greater than 2 m, the object was not detected. When the resolution of the object in the input image ( $1350 \times 1000$  pixel) was  $80 \times 140$  pixels or less, the precision of detection decreased by more than half, and when it was  $50 \times 80$  pixels or less, the object was not detected. For the medicine bottle, there was only one miss and no FP. Therefore, although the detection precision for the medicine bottle is high, a scenario should be created by considering the DR according to the distance. Accordingly, it was determined that it may be challenging to detect a medicine bottle in a real environment. In the experimental results, object detection was limited to the case of a medicine bottle placed on a dining table within 1 m from the system.



**Figure 11.** Bottle detection results based on distance: (a) 0.5 m, (b) 1 m, (c) 1.5 m, and (d) 2 m.

In the case of windows (Figure 12) and door, the Open image dataset was used. In the experimental environment, two windows and one door were present. When the distance from the system was 2 m or

more, the two windows were input into the system camera, and when the distance was less than 2 m, only one window was input. A DR of 12.5% and a precision of 86.57% were recorded when the distance from the system was 0.5 m. At 1 m, only one window was input, and a DR of 40.40% and a precision of 58.20% were recorded. The reason for the lower precision compared with the case of 0.5 m was the increased number of cases in which the window was misrecognized as a door. At 1.5 m, a DR of 79.31% and a precision of 83.29% were recorded, confirming that both the DR and the precision increased. When the distance from the system was 2 m or more, two windows were input, the number of objects to be detected in all frames accordingly increased two-fold. At 2 m, a DR of 86.23% and a precision of 100% were recorded, and for 5 m, a higher DR of 90.10% was recorded with 100% precision this was the highest DR. For the door, when the distance from the system was 5 m, a DR and precision of 100% were observed. When the distance from the system was 2 m or less, the door was not detected. For this reason, an additional experiment was conducted by increasing the distance to 3 m, and the results indicate that DR significantly decreased to 20.78%. For a large object, such as a door, the miss rate rather increased owing to occlusions of the object image when the distance between the object and the system decreased.



**Figure 12.** Window detection results based on distance: (a-1) 0.5 m TP, (a-2) 0.5 m miss, (b-1) 1 m TP, (b-2) 1 m miss, (c-1) 1.5 m TP, (c-2) 1.5 m FP, (d-1) 2 m TP, (d-2) 2 m miss, (e-1) 5 m TP, (e-2) 5 m miss.

In the case of the medicine bottle, the DR increased as the distance from the system decreased because the size of the actual object was small. In the case of windows and doors, as the distance between the system and the object increased, both the DR and precision increased. Consequently, the optimal location of the system can be chosen according to a real-world object. Furthermore, in general, windows and doors are located on walls inside a house. As the proposed system provides information and application in the real world through a projector, it is expected that a minimum distance of 3 m can be ensured between the system and windows/door. Furthermore, the YOLO deep-learning model can apply real-time object detection at a rate of 25 fps on average. Therefore, a seamless Deep-cARe system application can be provided to the user while a background real-time object detection process is run.

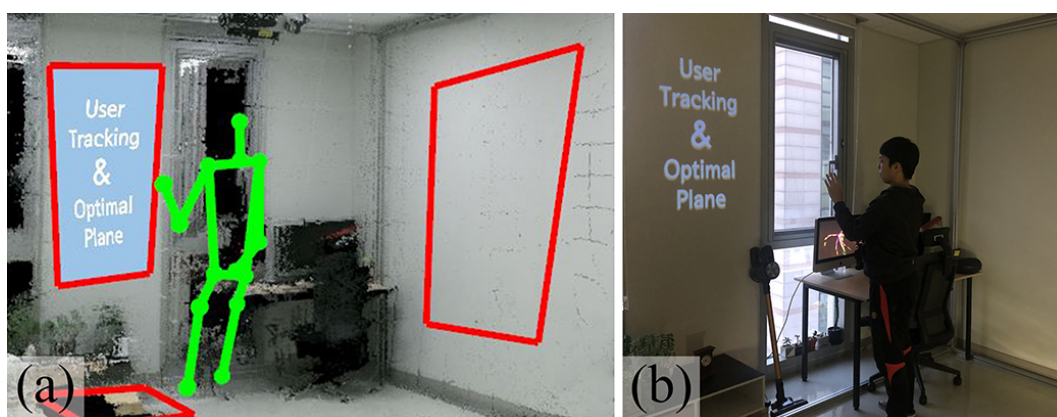
## 5. Deep-cARe System Application

The proportion of the world's older persons who are aged 80 years or over is projected to rise from 14% in 2015 to more than 20% in 2050 [1]. Therefore, we have designed and implemented the system with a simple application and an intuitive UI to provide functions to support the average 80 age elderly. The proposed system can prevent accidents by detecting an abnormal state of the elderly using the monitoring system. Furthermore, mental care support is provided, including physical and memory support activities.

### 5.1. Scenario Design

#### (a) Monitoring

The elderly can experience sudden injuries caused by accidental falls or tripping owing to the deterioration of their physical abilities. Occasionally, emergencies that may threaten their life can occur. This problem is more serious for those who live alone and spend considerable time indoors. There is a growing need to develop a monitoring system that can take the appropriate action in the event of an emergency inside a residence. Therefore, we developed a monitoring system that can identify such emergencies using a pan-tilt system and an RGB-depth camera. This system continuously tracks the user location using a camera after a 3D geometric environment around the user is reconstructed. The user location is essentially denoted by 3D-space coordinates using a 3D camera, which are used to augment information through the projection when the contents approach a registered plane (Figure 13). Furthermore, the system can recognize a fall accident or an emergency by identifying abnormal situations through the deep-learning pose estimation module. The situation of elderly people living alone often worsens because they cannot request medical support in case of an emergency. Therefore, the proposed system performs continuous monitoring of the location and state of the elderly while rotating the pan-tilt system and provides information to linked users so that they can prepare for emergencies. Accordingly, elderly users can receive proper support.



**Figure 13.** Monitoring scenario. (a) User in 3D map environment. (b) User in real environment.

(b) Smart IoT intercom

The elderly who often experience problems associated with long-term memory exhibit reduced ability to recognize faces [32]. In particular, they have difficulty identifying people quickly. The proposed system can resolve this problem. We propose an Internet of Things (IoT) application that can identify users in shown Figure 14. The scenario is as follows. Assuming that a visitor rings the doorbell, the elderly receives an information of the visitor. A visitor identification can be performed using the deep-learning face recognition module. Therefore, if the person on the screen is registered, his/her name is provided to the user. If the person is an unregistered previous visitor, "undefined" is displayed. However, if the person is a new visitor, "unknown" is displayed. In addition to classifying visitors and providing the relevant information, as shown in data management in the Figure 15, if someone is classified as "unknown," his/her input images are saved as an "undefined" class. For the "undefined" class, the face can be registered by saving the label information of the user. The system was designed to open or close the door by using a MUI (Figure 14c) or SUI (Figure 14-front wall). Therefore, for the elderly with discomfort in movement, physical support is provided by allowing remote control ability. Furthermore, long-term memory support is provided, and the elderly are alerted to unregistered visitors so that illicit activities may be prevented.

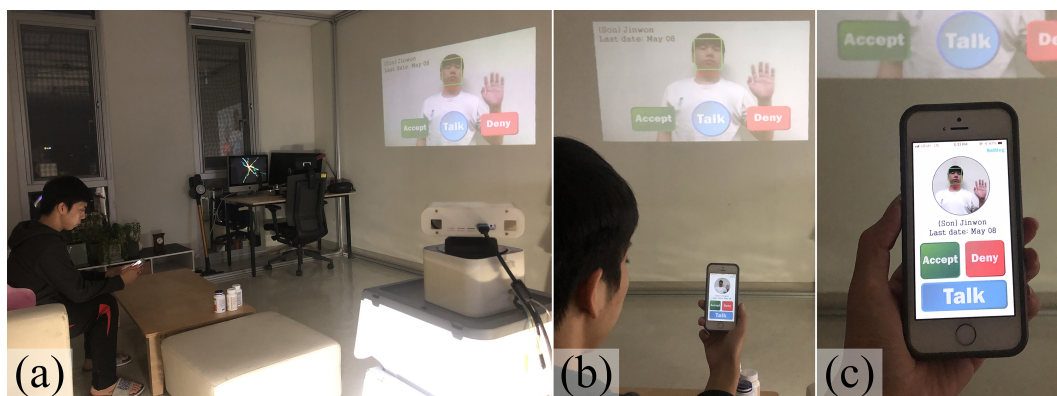


Figure 14. Smart IoT intercom scenario. (a) Intercom output. (b,c) Intercom UIs.

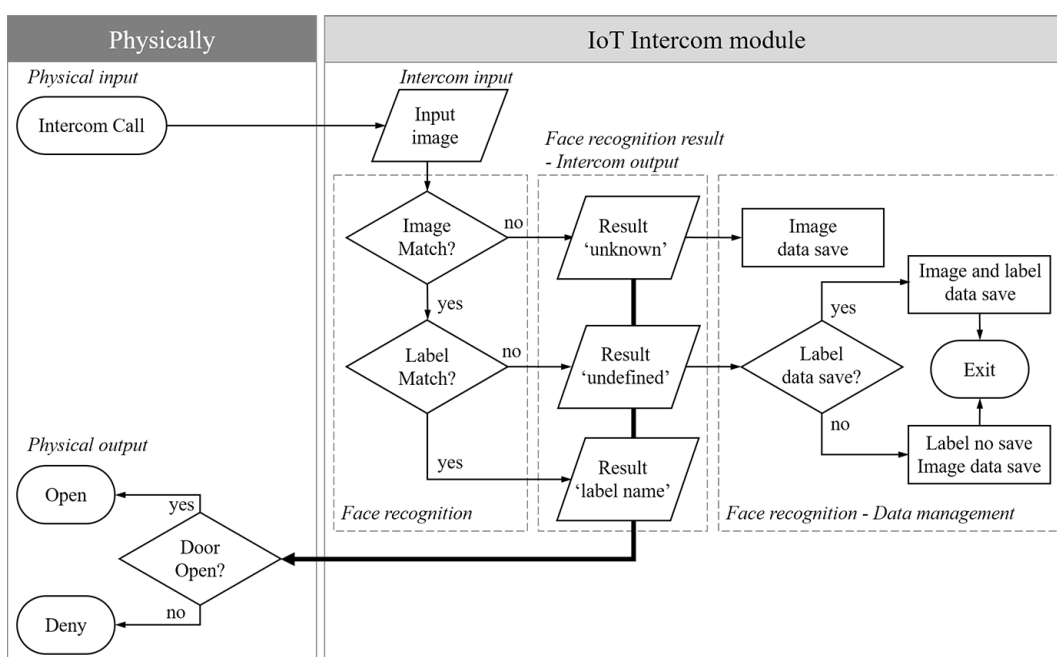


Figure 15. Smart IoT intercom scenario flow diagram.

(c) Daily Alarm

The elderly’s also experiences short-term and long-term memory problems, which result in forgetting important schedules or the state of the house. In addition to memory decline, the elderly also suffers from chronic diseases such as hypertension and diabetes. However, elderly patients can easily forget a dose or the time for taking a medicine. Therefore, their medication adherence (i.e., properly taking medicine according to the medical professional’s instructions) is poor. Medication notification can be provided through the proposed platform with no support devices, as in the case of smartphones and smart medicine. The purpose medication alert system intuitively and effectively notifies the time for taking medicines through audiovisual effects via a projector as shown in Figure 16. The UI for recognizing the medication and prompting its administration by using object detection is projected, and a sound alarm is provided to remind the user to take the medicine at the proper time.

In addition to medication notifications, additional daily alarm scenarios were created pertaining to information on real-world objects that can be frequently lost by the elderly in their daily lives. As shown in Figure 17, Deep-cARe system provide various types of information such as house internal condition and weather information. Door and window recognition can suggest



actions that can be taken before leaving. First, if the user is estimated to be near the door, the current behavior of the user is recognized as “leaving preparation state.” At this time, information such as the power stated of various devices (e.g., electric lights and electronic devices) and opening/closing of windows can be provided through projections near the door as shown in Figure 18. Consequently, the user can promptly receive information concerning the internal condition of the house and use a control center for the entire house before leaving. Furthermore, the system supports short-term memory by proposing clothes and accessories such as a hat and umbrella to prevent heatstroke. Moreover, through object detection, the elderly may be notified of the need to ventilate or open/close windows according to the weather using projections and sound alarms. This daily alarm induces the elderly to check the inside of the house and prevent accidents that can occur inside and outside the house.



Figure 16. Daily alarm scenario. (a) Object detection result. (b,c) Medication notification SUI.

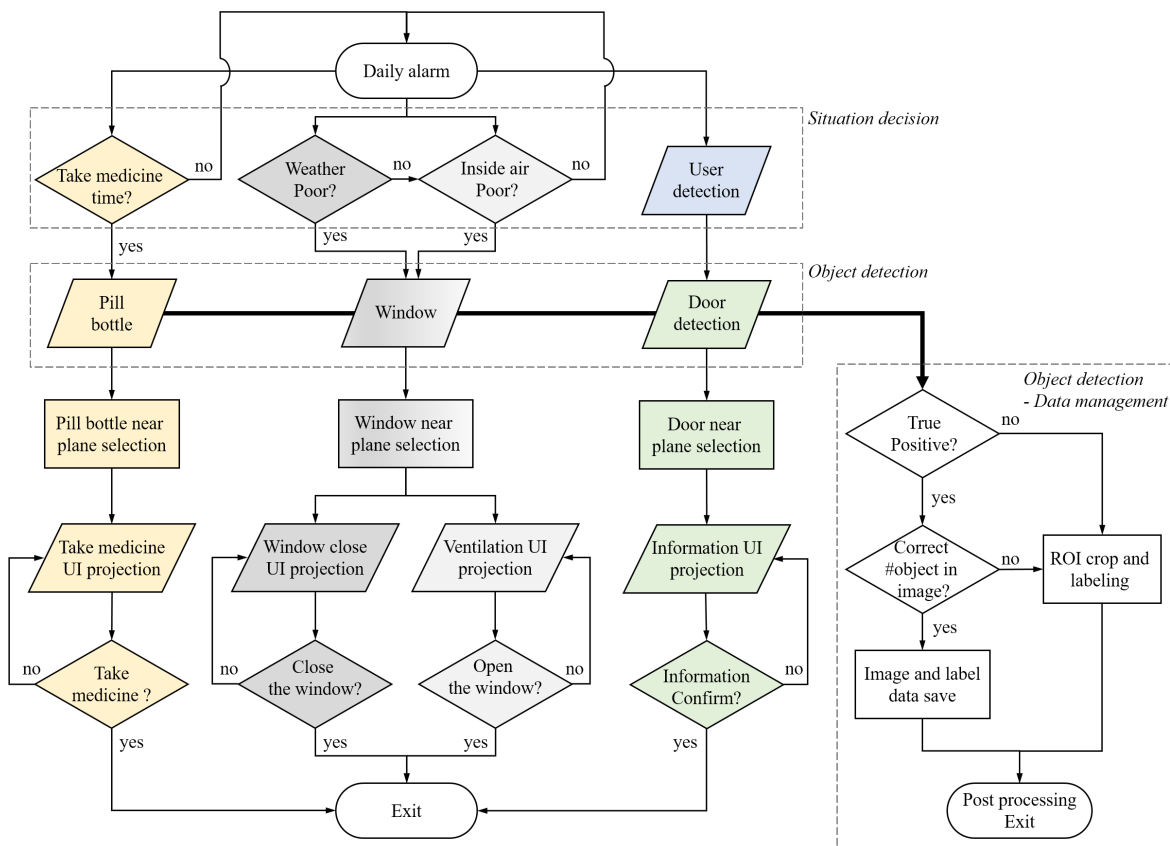
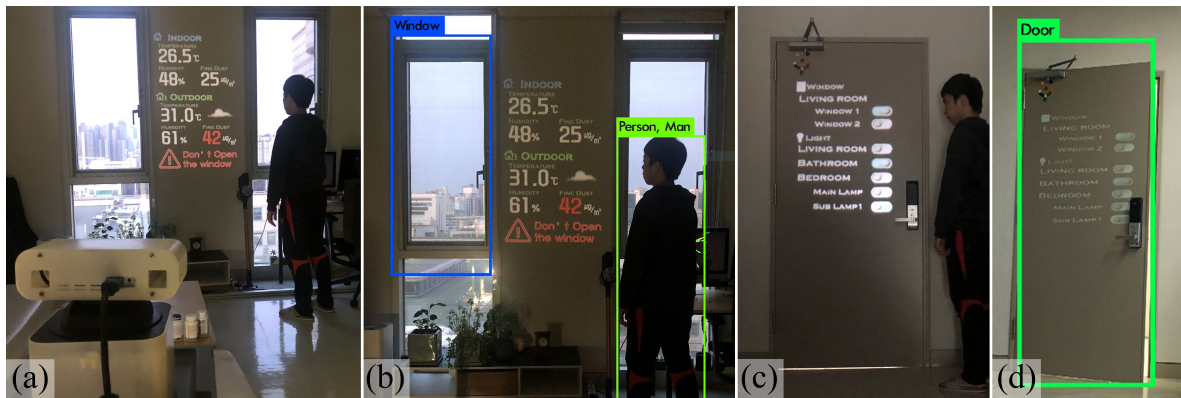


Figure 17. Daily alarm scenario flow diagram.



**Figure 18.** Weather condition and home control scenarios (a) Outdoor weather cast and indoor air state notification. (b) Object detection result of (a). (c) Home control UI. (d) Object detection result of (c).

## 5.2. User Interface and Interaction Design

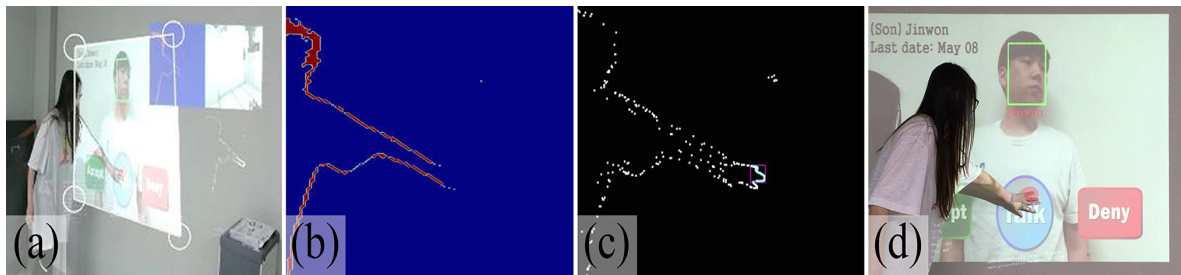
As the primary users of the proposed system are elderly people, it is critical to ensure a simple and intuitive interface and interaction. To this end, SUI and MUI that can mutually complement one another are provided instead of one interface. If the system and user are within the short distance, the SUI can provide an intuitive way of interaction. A convenient SUI is provided that allows directly touching the projected UI. In the case of the middle and long distance, the user uses the MUI so that they do not need to move and can interact with the system. A MUI with very low complexity is provided for the elderly who experience difficulty in using mobile devices.

### (a) Spatial user interface

To use a SUI in a PAR space where there is no touch sensor and panel, the touch interaction can be implemented using a depth-sensing camera. The formula of Wilson et al. [37] was applied to detect touch motions between the user and the PAR space. When the user touches the surface of the projection space, the depth pixel of the user appears nearer to the camera than the pixels of the projection space surface. In this case,  $d_{surface}$  in Equation (1) is the surface depth value of the projection space and stores the space surface depth information (Figure 19b). The elements that are considered to be in contact with the surface of the projection space are removed by setting the threshold of  $d_{max}$ . Pixels smaller than  $d_{min}$  have not contacted the surface of the projection space and other elements, and the contact of the user is not considered.

$$d_{surface} \leq d_{min} < d_{touch(x,y)} < d_{max} \quad (1)$$

When a change in depth information is detected in the area between the two thresholds, it is considered touch (Figure 19c). Interactions with the system are possible by directly touching the projected UI with no separate touch sensor using the space touch detection equation. However, in the case of the SUI, when the distance between the projected space and user is greater than a certain value, the user should move to the projected space. This can cause unnecessary movements for elderly people with movement difficulties. In this manner, the SUI can be used simply by directly touching the projected UI as shown in Figure 19d.



**Figure 19.** SUI (a) Input of color image and SUI touch area. (b) Input of depth data. (c) Sensing spatial touch point. (d) SUI touch result.

## (b) Mobile user interface

Understanding smartphone operation is no longer restricted to young people because smart devices such as smartphones and tablets are easily available and ubiquitous. Furthermore, IoT technology for controlling the inside of a house via smart devices has been developed. Nevertheless, the elderly still experience difficulty in using small icons or complex functions [38]. Instead of a complex interface that requires separate learning for mobile device operations, a simple mobile interface is provided. Even the elderly who experience discomfort in movement and are far from the system can actively use the system because a mobile interface is provided in accordance with the user's situation and application.

## 6. Conclusions and Future Work

### 6.1. Conclusions

This study proposed a pervasive Deep-cARe system that provides care to the elderly by constructing a PAR environment for the real-world space in which the elderly live. The Deep-cARe system is combined with a deep-learning framework. We constructed projection-cam hardware with a 360° pan-tilt mechanism that can cover all spaces around the user. A PAR space that can provide virtual information, content, and application in the real-world space can be obtained by reconstructing the real-world space in a 3D map. By combining deep-learning pose estimation, including face and object-recognition technologies, the Deep-cARe system can be applied to a complex real-world environment with no prior definition. This paper describes the actual content of implementation and we intend to demonstrate through experiments that the globally learned deep learning with PAR system can support apps for elderly care. Experiments were conducted to evaluate the performance of deep-learning pose estimation, face recognition, and object detection in the actual environment. For pose estimation results of "stand" state, the average precision of all joints was 86.98%. "seat" state, overall precision of 87.47%. The lower body joints were not estimated at "hide" state and the face and upper body were estimated at "lie" state. The precision of the face pose is particularly high by all state of pose estimation. In case face recognition, the average DR was 74.84% and the precision was 78.72% for face recognition results of whole class. However, for face occlusions, the average DR was 46.83%. It was confirmed that the face recognition can be performed properly if the face occlusion situation is not occurring frequently. By object detection experiment results, The DR increased as the distance from the system decreased for a small object (bottle). For a large object such as window and door, the miss rate rather increased when the distance between the object and the system decreased. These scenarios include safety application, the cognitive abilities. According to the scenarios, MUI and SUI were designed and implemented, and their functions supplement one another depending on the situation. We are currently in the phase of designing and building the PAR and deep-learning fusion platform as the first step of our research. Furthermore, we should verify that the proposed system can achieve a more effective elderly care system than traditional systems by actually providing it to the elderly living.

## 6.2. Limitation and Future Work

### (a) Context-aware accuracy

The proposed system improved the ease of installation and use. However, it is difficult to perform context-aware operations according to the elderly's behavior and environment using a single camera. In the future, an environment for highly accurate context awareness of the elderly's behavior inside a house should be constructed by installing multiple cameras and sensors.

### (b) Connection with IoT device

This study described an IoT intercom for remote door opening and closing. However, home appliances include more functions and their control methods are becoming increasingly complex. Therefore, they can be difficult for the elderly to use. The elderly may find it physically burdensome to open or close windows and doors. They could be simply and intuitively controlled through an interaction between various IoT devices and the proposed platform.

### (c) Application of broad deep-learning technologies

An AR environment for expanding the functions of the elderly and supporting their life can be provided by applying technologies with development potential in addition to the deep-learning technologies applied in this study. It is also possible to analyze correlations with the diseases that afflict individuals by analyzing the behavior of the elderly through activity recognition, which can be considered an extension of pose estimation. Furthermore, emotional care and physical support can be provided through feedback and therapy, in line with the user's current emotion, using emotion recognition technology.

### (d) System expansion

When a care system is designed, usability and user experience are important factors that must be considered. User experience can be enhanced by analyzing usability through a study involving multiple users. Therefore, the application of the proposed system can be expanded to sanatoriums, in addition to homes.

**Author Contributions:** This paper is a result of the full collaboration of all the authors; conceptualization, Y.J.P., H.R., T.-D.H.; implementation, Y.J.P., H.R.; validation, N.K.L., T.-D.H.; formal analysis, Y.J.P., H.R.; writing, Y.J.P.

**Funding:** The National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No.NRF-2018R1A2A1A05078628).

**Acknowledgments:** This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP), Department of Computer Science at Yonsei University.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. McNicoll, G. World Population Ageing 1950–2050. *Popul. Dev. Rev.* **2002**, *28*, 814–816.
2. Kim, K. Policy Considerations for the Living Conditions of Older Koreans. *Health Welf. Forum* **2015**, *223*, 68–78.
3. Jacoby, R.; Oppenheimer, C. *Psychiatry in the Elderly*; Oxford University Press: New York, NY, USA, 1991.
4. Lee, S.J.; Kwon, H.J.; Lee, Y.S.; Min, B.A. Analysis of Universal Design Bathroom and Its Products Supporting Aging in Place of the Elderly. *Archit. Inst. Korea* **2007**, *23*, 125–134.
5. Rashidi, P.; Mihailidis, A. A survey on ambient-assisted living tools for older adults. *IEEE J. Biomed. Health Inform.* **2012**, *17*, 579–590. [[CrossRef](#)]
6. Uddin, M.; Khaksar, W.; Torresen, J. Ambient sensors for elderly care and independent living: A survey. *Sensors* **2018**, *18*, 2027. [[CrossRef](#)] [[PubMed](#)]
7. Kawamoto, A.L.S.; da Silva, F.S.C. Depth-sensor applications for the elderly: A viable option to promote a better quality of life. *IEEE Consum. Electron. Mag.* **2017**, *7*, 47–56. [[CrossRef](#)]

8. Lu, N.; Wu, Y.; Feng, L.; Song, J. Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 314–323. [[CrossRef](#)] [[PubMed](#)]
9. Zhu, W.; Lan, C.; Xing, J.; Zeng, W.; Li, Y.; Shen, L.; Xie, X. Co-Occurrence Feature Learning for Skeleton Based Action Recognition Using Regularized Deep LSTM Networks. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
10. Khanal, S.; Reis, A.; Barroso, J.; Filipe, V. Using Emotion Recognition in Intelligent Interface Design for Elderly Care. In Proceedings of the World Conference on Information Systems and Technologies, Naples, Italy, 27–29 March 2018; pp. 240–247.
11. Mezgec, S.; Koroušić Seljak, B. Nutrinet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients* **2017**, *9*, 657. [[CrossRef](#)] [[PubMed](#)]
12. Wolf, D.; Besserer, D.; Sejunaite, K.; Riepe, M.; Rukzio, E. cARe: An Augmented Reality Support System for Dementia Patients. In Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings, Berlin, Germany, 14–17 October 2018; pp. 42–44.
13. Aruanno, B.; Garzotto, F. MemHolo: Mixed reality experiences for subjects with Alzheimer’s disease. *Multimed. Tools Appl.* **2019**, *78*, 1–21. [[CrossRef](#)]
14. Vovk, A.; Patel, A.; Chan, D. Augmented Reality for Early Alzheimer’s Disease Diagnosis. In Proceedings of the Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; p. LBW0212.
15. Ingeson, M.; Blusi, M.; Nieves, J.C. Microsoft HoloLens-A mHealth Solution for Medication Adherence. In Proceedings of the International Workshop on Artificial Intelligence in Health, Stockholm, Sweden, 13–14 July 2018; pp. 99–115.
16. Hockett, P.; Ingleby, T. Augmented reality with HoloLens: Experiential architectures embedded in the real world. *arXiv* **2016**, arXiv:1610.04281
17. Lera, F.J.; Rodríguez, V.; Rodríguez, C.; Matellán, V. Augmented reality in robotic assistance for the elderly. In *International Technology Robotics Applications*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 3–11.
18. Chae, S.; Yang, Y.; Choi, H.; Kim, I.J.; Byun, J.; Jo, J.; Han, T.D. Smart Advisor: Real-Time Information Provider with Mobile Augmented Reality. In Proceedings of the 2016 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 7–11 January 2016; pp. 97–98.
19. Schlomann, A.; Rasche, P.; Seifert, A.; Schäfer, K.; Wille, M.; Bröhl, C.; Theis, S.; Mertens, A. Augmented Reality Games for Health Promotion in Old Age. In *Augmented Reality Games II*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 159–177.
20. Bianco, M.L.; Pedell, S.; Renda, G. Augmented Reality and Home Modifications: A Tool to Empower Older Adults in Fall Prevention. In Proceedings of the 28th Australian Conference on Computer-Human Interaction, Launceston, Australia, 29 November–2 December 2016; pp. 499–507.
21. Yamamoto, G.; Hyry, J.; Krichenbauer, M.; Taketomi, T.; Sandor, C.; Kato, H.; Pulli, P. A User Interface Design for the Elderly Using a Projection Tabletop System. In Proceedings of the 2015 3rd IEEE VR International Workshop on Virtual and Augmented Assistive Technology (VAAT), Arles, France, 23 March 2015; pp. 29–32.
22. Pizzagalli, S.; Spoladore, D.; Arlati, S.; Sacco, M.; Greci, L. HIC: An Interactive and Ubiquitous Home Controller System for the Smart Home. In Proceedings of the 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH), Vienna, Austria, 16–18 May 2018; pp. 1–6.
23. Vogiatzaki, E.; Krukowski, A. Maintaining Mental Wellbeing of Elderly at Home. In *Enhanced Living Environments*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 177–209.
24. Petsani, D.; Kostantinidis, E.I.; Diaz-Orueta, U.; Hopper, L.; Bamidis, P.D. Extending Exergame-Based Physical Activity for Older Adults: The e-Coaching Approach for Increased Adherence. In Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health, Las Vegas, NV, USA, 15–20 July 2018; pp. 108–125.
25. Montemerlo, M.; Thrun, S.; Koller, D.; Wegbreit, B. FastSLAM 2.0: An Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges. In Proceedings of the IJCAI, Acapulco, Mexico, 9–15 August 2003; pp. 1151–1156.
26. Leutenegger, S.; Chli, M.; Siegwart, R. BRISK: Binary Robust Invariant Scalable Keypoints. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2548–2555.

27. Whelan, T.; Kaess, M.; Johannsson, H.; Fallon, M.; Leonard, J.J.; McDonald, J. Real-time large-scale dense RGB-D SLAM with volumetric fusion. *Int. J. Robot. Res.* **2015**, *34*, 598–626. [CrossRef]
28. Xu, L.; Oja, E.; Kultanen, P. A new curve detection method: randomized Hough transform (RHT). *Pattern Recognit. Lett.* **1990**, *11*, 331–338. [CrossRef]
29. Kendall, A.; Grimes, M.; Cipolla, R. PoseNet: A Convolutional Network for Real-Time 6-dof Camera Relocalization. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2938–2946.
30. Zhao, R.; Ali, H.; Van der Smagt, P. Two-Stream RNN/CNN for Action Recognition in 3D Videos. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 4260–4267.
31. Hai, P.T.; Kha, H.H. An Efficient Star Skeleton Extraction for Human Action Recognition Using Hidden Markov Models. In Proceedings of the 2016 IEEE Sixth International Conference on Communications and Electronics (ICCE), Ha Long, Vietnam, 27–29 July 2016; pp. 351–356.
32. Kawagoe, T.; Matsushita, M.; Hashimoto, M.; Ikeda, M.; Sekiyama, K. Face-specific memory deficits and changes in eye scanning patterns among patients with amnesic mild cognitive impairment. *Sci. Rep.* **2017**, *7*, 14344. [CrossRef]
33. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
34. Redmon, J. Darknet: Open Source Neural Networks in C. 2013–2016. Available online: <http://pjreddie.com/darknet/> (accessed on 30 June 2016).
35. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2014; pp. 740–755.
36. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2d Human Pose Estimation: New Benchmark and State of the Art Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3686–3693.
37. Wilson, A.D. Using a Depth Camera as a Touch Sensor. In Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, Saarbrücken, Germany, 7–10 November 2010; pp. 69–72.
38. Mohadisdudis, H.M.; Ali, N.M. A Study of Smartphone Usage and Barriers Among the Elderly. In Proceedings of the 2014 3rd International Conference on User Science and Engineering (i-USEr), Shah Alam, Malaysia, 2–5 September 2014; pp. 109–114.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).