A Face Recognition System for Automated Door Opening with parallel Health Status Validation Using the Kinect v2

A. Ogawa[a] , A. Mita[a], C. Georgoulas[b] and T. Bock[b]

[a]Department of System Design Engineering, Keio University, Kanagawa, Japan
[b]Chair for Building Realization and Robotics, Technical University of Munich, Germany
E-mail: ami_ogawa@keio.jp, mita@keio.jp, christos.georgoulas@tum.de, thomas.bock@tum.de

Abstract

Nowadays it is said that extending the healthy life expectancy is important in the aging society target group. Of course it is important for every single-person household to be regularly checked upon their safety, but especially for elderly people the importance is higher at the point of healthy living, because this particular target group comprises a higher accidents rate. Therefore, the authors here propose the design and implementation of an unobtrusive vision system for single-person households, particularly for elderly people, based on the Microsoft Kinect v2. The entrance area of the home environment is here considered, where the monitoring system is activated upon while the user approaches, in order to first detect and calculate the user's fatigue levels, as well as to secondly identify the user by facial recognition, in order to actuate a door opening mechanism. The proposed system algorithm is divided into two distinct phases. The first phase concerns the acquisition of the user knee position while walking up a staircase in order to calculate fatigue. The second phase comprises the face recognition based door opening method. The proposed system has been successfully tested, and it could undoubtedly comprise an unobtrusive health status validation and automated door opening solution for elderly people in single-person households.

Keywords –

Computer Vision, Kinect v2, Face Recognition, Fatigue Detection, Ambient Assisted Living

1 Introduction

Nowadays the number of single-person households is increasing and it is expected to rapidly grow within the following years [1,2]. In Norway Finland and Denmark the single person households already reached more than 35%. In Japan this percentage is currently at 33%. One of the concerning things is that the single-person households have a higher risk of exacerbation of the accidents which happen in living spaces than the others. It is because in the case of the single-person households, normally there are no other people present in their living spaces, so they cannot retrieve any assistance or care in case of an accident.

This is a seriously concern, especially for elderly people. Elderly people are affected heavily because they have less ability of recovering. For example, if they fall down, many of them might not be able to walk and this could psychologically affect them. Therefore, an unobtrusive monitoring system could increase their safety.

To decrease the risk of accidents in living space for single-person households, the detection of the resident's physical and mental information in real-time is required, in order to be able to recognize risk situations. Considering the time flow of the monitoring system in the living space, the sensing is activated when the resident enters the house. Therefore, it is more effective to position the sensing system at the entrance part so that we can retrieve as early as possible the resident's initial condition.

The authors eventually aimed to predict the resident's physical and mental fatigue. As the first step of it, in this research, it was suggested the combined system which consists of the knee position acquisition of stair walking and face recognition based door opening. The knee position acquisition is expected to be extended to get the feet joint parameters for physical fatigue evaluation. And face recognition is expected to be a tool for aspect prediction for mental fatigue evaluation. In this study, we used the Kinect v2 which provides RGB-color, IR-depth and IR images with 1920x1080, 512x424 and 512x424 spatial pixel resolution respectively, and can detect the human body and track 25 joints without putting any markers [3]. The proposed system has been implemented and realized in the experimental ambient assisted living laboratory of the authors at the Technical University of Munich.

2 Background

There are various attempts addressing smart home environments embedded with many sensors and actuators inside walls, floors, ceilings, etc, which retrieve information about the environment accumulate into a sort of live log, to optimally control smart devices, and to make residents comfortable [4-7]. However, all these attempts also impose limitations due to the necessary and expensive reconstruction costs in order to install and embed all sensors and actuators. In some situations large scale reconstruction is required or even to move to a brand new smart home construction. Moreover, even if we can live in such smart house once, we have to change the sensors and devices often, because nowadays the technology advances with higher rates compared to a buildings lifecycle. Thus, instead of focusing on a smart wall, or smart floor, or smart ceiling approach, the authors chose the so-called "Terminal" approach[8].

The "Terminal" comprises a modular furniture containing specific functions and services depending on the room or life center it is installed. It can be said a kind of furniture with several sensors in simple words, so that it is easy to install into existing buildings. This concept saves us from renovating or reconstructing the whole house. It is designed and manufactured by combining standard products to save the cost of production. Moreover due to the fact that it is "modular", it can straightforwardly be adapted to any kind of house. Every house has its unique features and arrangement. Additionally every user might require a different level of intelligence (considering the terminal sensors and actuators), so modularity here allows for user customization.

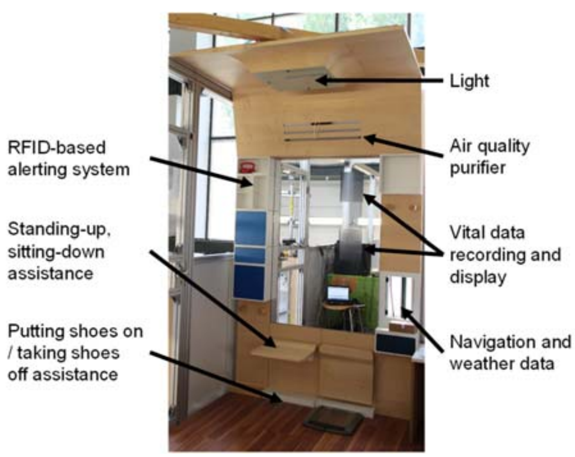


Figure 1. The entrance "Terminal", LISA smart Wall [8]

An example of such terminal can be seen in Figure 1, which was developed under the research project LISA, investigating the possibilities to embed mechatronic, assistive functions and services into compact wall "terminal" elements thereby enabling autonomous and independent living upon performing Activities of Daily Living (ADLs) by means of generated structured environments and robotic micro-rooms (RmRs) 8. The proposed system in this paper comprises an add-on for the terminal shown in Figure 1, since its output can be fed to the terminal onboard screens, to display the resulting user fatigue condition at the time, and provide with necessary advices. Also, it is possible to control the lighting and air conditioning devices on the L.I.S.A. wall based on the results.

3 Proposed Implementation

The proposed system has been implemented and realized in the experimental ambient assisted living laboratory of the authors at the Technical University of Munich. A full scale home environment has been built, comprising entrance space, living room, bed room and bathroom, kitchen. There are eight steps towards the entrance door, and the entrance door is simply composed of steel frames (Figure 2). The Kinect was used to acquire RGB, depth, and also IR consecutive images as an animated film. A 3D models was created by combining these data. Moreover, Kinect v2 (Figure 3) has a skeleton tracking function, face tracking function, and microphone arrays. In the proposed implementation, the skeleton tracking and depth data were used for knee position acquisition, and face tracking and IR images for face recognition respectively (Figure 4).

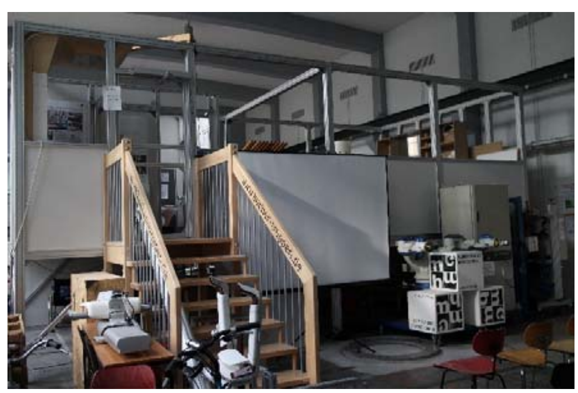


Figure 2. Experimental ambient assisted living laboratory

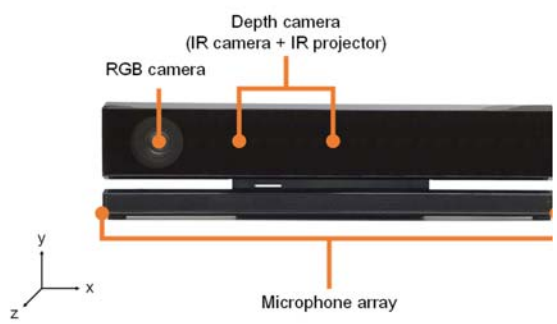


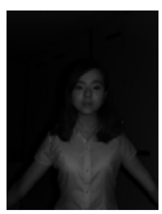
Figure 3. MS Kinect v2

3.1 Proposed Knee Position Acquisition Method

This knee position acquisition while walking up the stairs was proposed as a prediction of the user fatigue level upon entering the home environment. It is meaningful to know the user condition upon arriving home. If the user fatigue level is high, the system can warn to notice upon the condition, in order to prevent any potential accidents due to it. Kinect v2 can detect up to 6 people's bodies. It can not only detect human body, but also mark the 25 different joint points. The available range of this function is limited in 4.5m from Kinect, but the entrance approach of experimental house was within the range, so the efficiency of the system was not questioned.

a. RGB image

b. IR-Depth image



c. IR image

d. Skeleton tracking

e. Face tracking

Figure 4. The main function of Kinect v2

Firstly, the knee position while walking up the stairs was detected and retrieved using the skeleton tracking function directly. The visualized joint position marks are shown in Figure 5. According to the original knee position, it is evident that the mark does not reside with the exact knee position. Then knee position correction was considered based on the following two facts about the features of stair walking.

- Fact 1: The true knee position is always between the knee position and the ankle position which are given by Kinect v2 (Figure 6).
- Fact 2: The true knee position is always the forefront of forward movement in the part of the foot (Figure 7).

Figure 5. The original joint position of Kinect v2

Then the knee position correction algorithm was implemented which is based on the knee and anckle position provided by the Kinect. Firstly, the 3-D value of the knee and ankle position is obtained. Then the straight line from the knee to ankle position is calculated according to equations (1), (2) and (3). At this time, only y-z plane is considered:

$$y = mz + n \tag{1}$$

$$m = \frac{y_{Knee} - y_{Ankle}}{z_{Knee} - z_{Ankle}} \tag{2}$$

$$n = y_{Knee} - \frac{(y_{Knee} - y_{Ankle})z_{Knee}}{z_{Knee} - z_{Ankle}} \tag{3}$$

After the points along the line on the x-y plane are examined one by one. Then the distance between the depth, z value, of each point and the straight line which connects knee and ankle position in the y-z plane is calculated. The equation of the distance between line and each position is expressed in equation (4):

$$d = \frac{|y_i - mz_i - n|}{\sqrt{1 + m^2}} \tag{4}$$

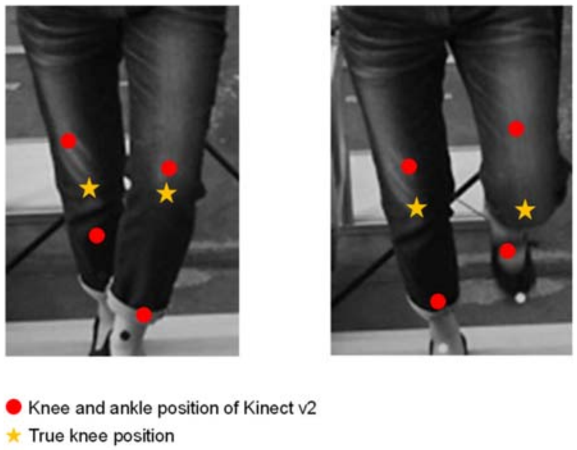


Figure 6. True knee position is always between the original position of knee and ankle

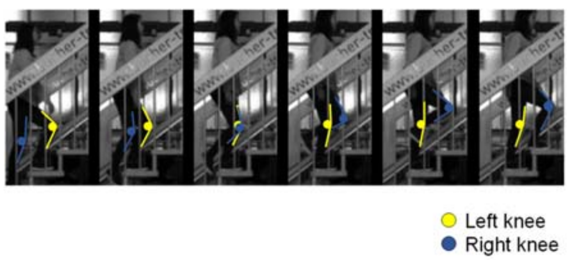


Figure 7. True knee position is always the fore-front of forward movement in the part of the foot

Finally, we can detect the point which has the maximum distance from the straight line as the true knee position (Figure 8).

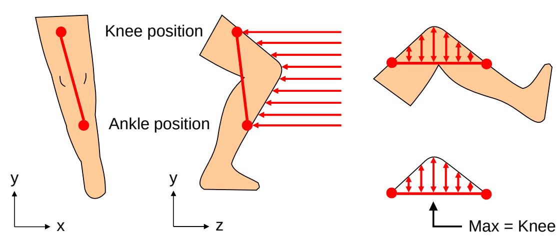


Figure 8. The proposed knee position detection algorithm

Applying this method, the accuracy of knee position was seriously improved. However, it is not always accurate because it is not sure that we can always get the knee and ankle position by Kinect v2. For this reason another 2 post-processing steps to accurately calculate the true knee position were developed. In case the calculated value was incorrect, i.e. extremely small, extremely large, or out of range, the value was rejected by filtering. In particular, the value will be rejected when the z value difference of calculating knee and spine base is more than 1000mm. Sometimes we cannot have the value of the specific position. Especially those kind of error is happening with hands and feet and the accuracy is unstable. On the other hand, the most stable joint position is spine base. Therefore, we used the difference with spine base. And the true knee will be the smallest value selected in the area of the rectangle which consist of knee and ankle x and y position of Kinect v2 instead (Figure 9).

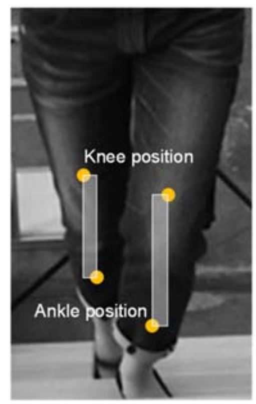


Figure 9. The range of knee detection for 1rst post processing step

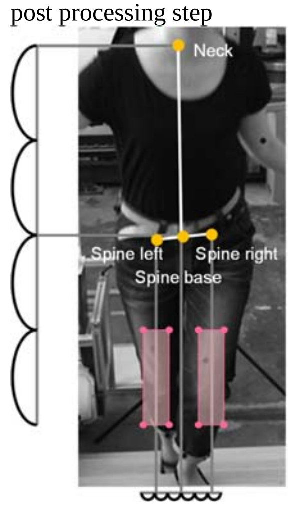


Figure 10. The range of knee detection of 2nd post processing correction

Finally the last post-processing step was applied in the case of no or incorrect value of knee and ankle position given by the Kinect v2. Same as the previous step, the value will be recognized as an incorrect value when the z value difference of the knee or ankle and spine base is more than 1000mm. The program is written to find the true knee position by using spine base, spine left, spine right and neck to define the range of the rectangle, and extract the minimum depth value inside of this range. In particular, the definition of the range of rectangle is shown in Figure 10. The y distance is defined as half of the y distance between the neck and the spine base. And the x distance is defined as the x distance between spine base and spine left or right centering on spine left or right. The total algorithm flow is shown in Figure 11.

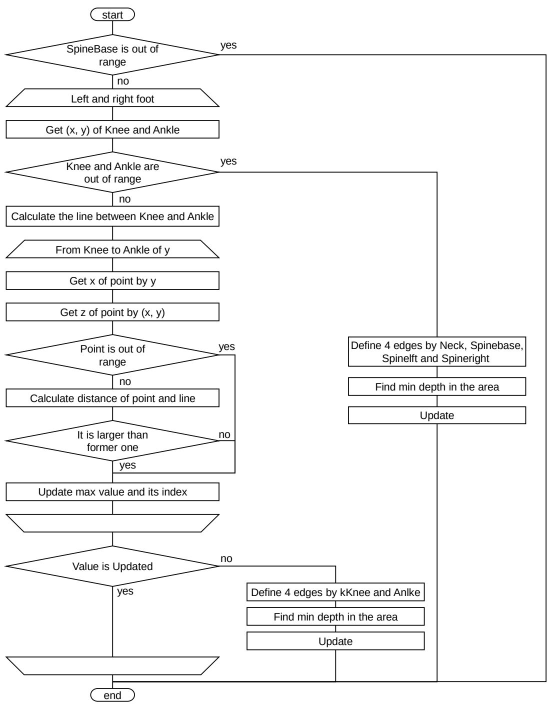


Figure 11. Flowchart of the proposed algorithm for knee position detection

3.2 Proposed Face Recognition Method

The face recognition based door opening system was placed on the entrance door of the experimental house. Thus, it will be held after the knee position acquisition. Normally depth data should be used for sensing because of the privacy protection in living space where is a very personal place and hard to accept putting cameras inside. However, this face recognition system is supposed to be used in front of the entrance, which means outside of the house just like an intercom, so that there is no need to consider about resident's privacy because outside of the house is no longer public space. Therefore, we can use any of the RBG, IR-depth or IR images. In the proposed approach only RGB and IR images were used for the face recognition function. Firstly, RGB images were used, but in the feature extraction phase, due to the fact that the features in the images depend on the lighting condition, the accuracy was unstable (Figure 12). Therefore, IR images were used instead. The proposed face recognition flow consists of four phases: 1) Face detection, 2) Feature extraction, 3) Read of database, and 4) Identification of the person.

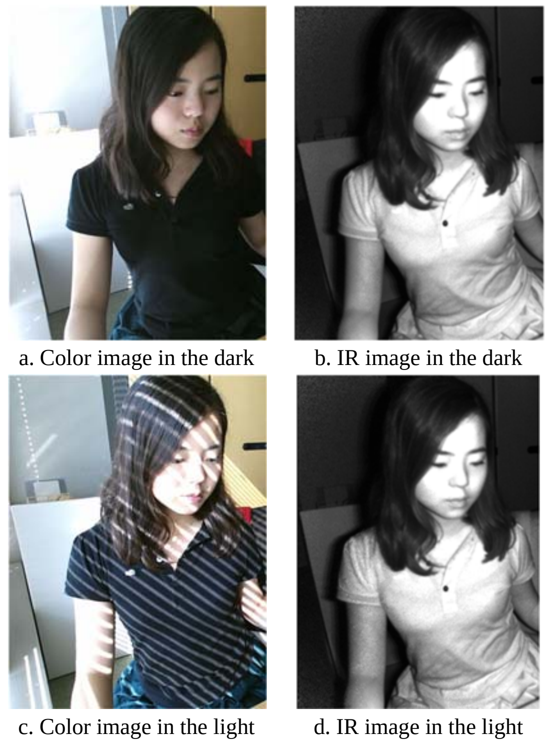


Figure 12. RGB images are affected by the light condition but IR images are not

3.2.1 Face Detection

As we mentioned before, Kinect v2 has the face tracking function which actually works based on the skeleton tracking function, i.e. it can detect the face automatically. Therefore, we used this function directly.

3.2.2 Feature Extraction

Here the Local Binary Pattern Histogram [9, 10] was used as the feature extraction method of the face recognition. Local Binary Pattern Histogram is one of the local feature descriptor and is widely used for face recognition [11]. Also, it is evaluated as a balanced feature extraction method at the point of accuracy and processing speed [12]. Therefore, the authors used LBPHFaceRecognizer provided by OpenCV.

3.2.3 Image Retrieval Database

Several JPG images were obtained of various subjects for the generation of the database and an Excel file was created listing the image file names. The size of the database was also concerned. The larger the number of images in the database, the higher the performance of the face detection algorithm, but the processing speed is slower, and vice versa. Eventually we collected three subject's images in the database.

Figure 13. Example of JPG images in the database

3.2.4 Person Identification

If the input image identified specific person who were saved in database previously, it means face recognition is done. The threshold of the classification after a series of experiments was empirically set to 50.0. If the distance from the feature value of the input image to any prototypes is larger than the threshold, the input data is rejected and the person is recognized as a "Stranger".

3.3 Proposed Door Opening System

Using and Arduino Uno microcontroller board interfaced with the door opening mechanism (Figure 14) as well as with the Kinect, following a successful face recognition result, the result could be used to enable a signal on the microcontrollers output to actuate the door opening mechanism. When the user's face is recognized as the resident, a logical '1' is enabled on a specific pin on the Arduino GPIO, which is used as a signal to drive the door opening mechanism. Also a green LED is lit for visual response. On the other hand, when the user's face is not recognize as any of the residents stored in the database, meaning the user is recognized as a stranger, a logical '0' is enabled on the GPIO pin. Additionally a Red LED is lit accordingly for visual purposes. With a logic '0' the door opening mechanism remains idle, i.e. the door remains closed. The door operating system is programmed to close the door several seconds after the door opening.

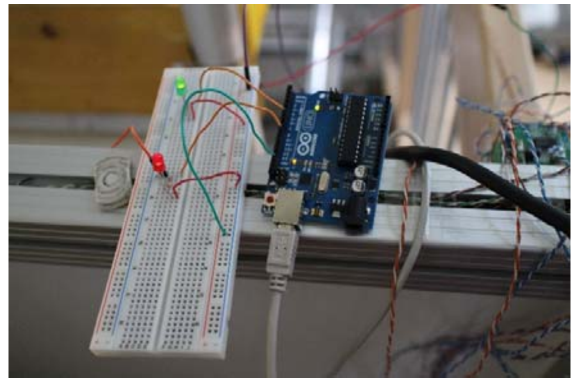


Figure 14. ARDUINO Uno Board

4 Testing

The system performance was tested following the installation of the Kinect sensor in the entrance of the flat (Figure 15). The Kinect was set at an angle of 34.3 degrees. Both subsystems, knee position acquisition in stair walking and face recognition based door opening, were implemented and developed in C++. The testing followed the steps below:

- When the Kinect detects a subject, the knee position acquisition of stair walking is initiated. The calculated data is accumulated temporarily.
- When the user terminates the walking sequence i.e. arrives at doorstep, the knee position acquisition is interrupted and the face recognition peocess starts.
- If the subject is recognized as a resident within the database the door automatically opens.

The trigger to switch from stair walking acquisition to face recognition is the distance between subject and the Kinect. The distance is defined as the depth value of spine base position which is given by Kinect and it was 1150 mm.

5 Conclusions

In this paper, a combined system is proposed offering knee position acquisition while walking on stairs as well as face recognition based automatic door opening. These combined system aim to predict both the mental and the physical fatigue of the user upon returning to the home environment.

In the knee position acquisition phase, a novel knee position correction algorithm is proposed. According to the acquired results, the accuracy is undoubtedly improved. In the face recognition phase, images depicting the faces of three users are stored in a database. The number of images per user, as well as the number of users has been considered for the trade off between face recognition accuracy and processing speed. The system was installed and tested under realistic 1:1 scale in order to prove its possibility for realization. The position of the Kinect sensor in the experimental setup was extensively tested.

Currently the authors are conducting accuracy verification experiments for both subsystems. As future development, the face recognition phase can be expanded to detect the mental fatigue level according to the facial expressions of the user. It is expected that the Kinect v2 detailed face tracking function can be exploited towards this goal. Considering the knee position acquisition phase, it is expected to first predict the fatigue levels by comparing the acquired data and accumulated data within the database, and then enable warnings and advices depending on the calculated fatigue level.

The recognition can be used not only to actuate the door opening mechanism, but also to access to the user personal database including gait data for this study. This can enable a personalized actuated environment since lighting, air conditioning, aroma diffuser, music player, etc can be automatically controlled for the specific resident and his / her condition which is predicted from the fatigue level etc.

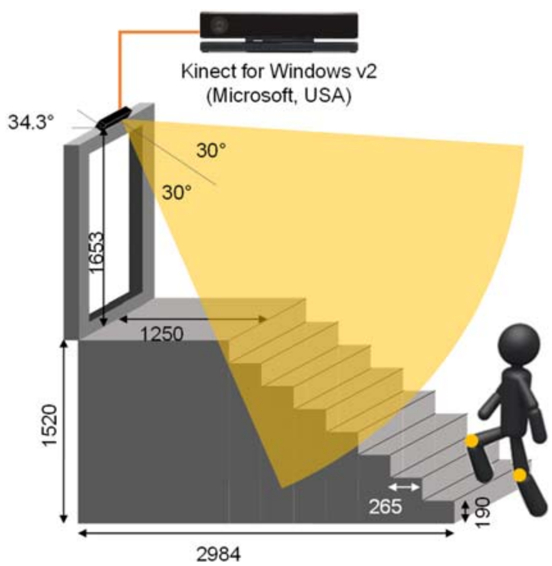
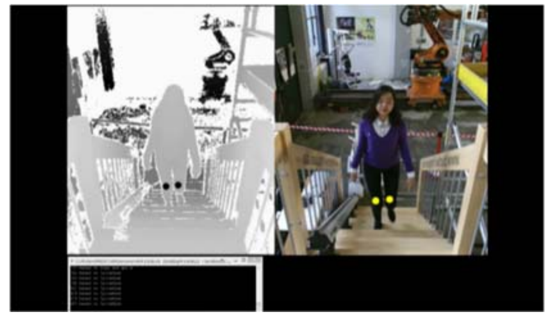
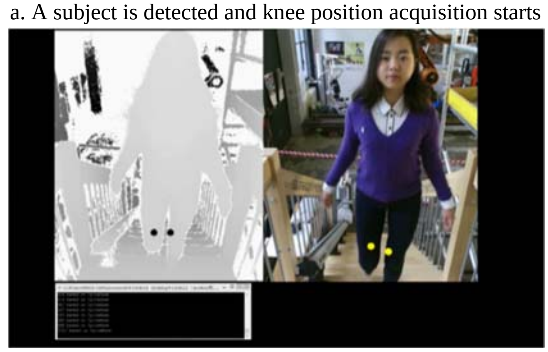


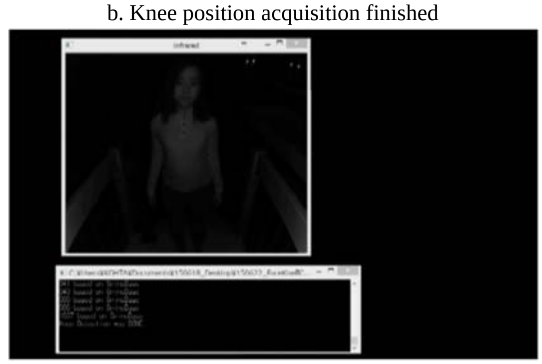
Figure 15. System Setup Arrangement



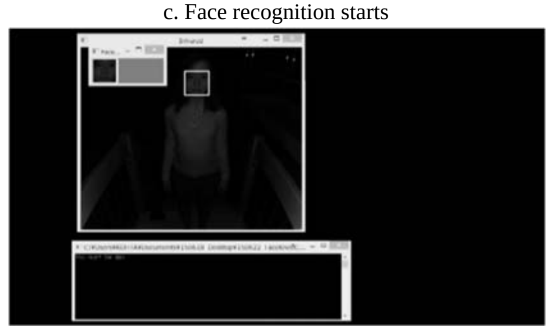
a. A subject is detected and knee position acquisition starts



b. Knee position acquisition finished



c. Face recognition starts



d. The subject is recognized as a resident, and the door will open

Figure 16. Execution screen of the testing sequence

References

[1] Yeung W.J.J. and Cheung A.K.L. Living alone: One-person households in Asia. Demographic Research, 32(40):1099-1112, 2015.

[2] Podhisita C. and Xenos P. Living alone in South and Southeast Asia: An analysis of census data. Demographic Research, 32(41):1113, 2015.

[3] Amon C and Fuhrmann F. Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1and v2. In Proceedings of the 6th Congress of the Alps Adria Acoustics Associa-tion, Graz, Austria, 2014.

[4] Waseda Univ, Wabot house, On-line: http://www.wabot-house.waseda.ac.jp/html/e-top.htm, Accessed: 15.03.2016.

[5] Murakami K., Hasegawa T., Kimuro Y., Karazume R. A structured environment with sensor networks for intelligent robots. In Proceedings of IEEE Sensors, pages 705-708, Lecce, Italy.

[6] Initlle S., Larson K., Tapia E.M., Beaudin J., Kaushik P., Nawyn J., Rockinson R. Using a live-in laboratory for ubiquitous computing research. In Proceeding of the 4th International Conference on Pervasive Computing, pages 349–365, Dublin, Ireland, 2006.

[7] Larson K and Stephen I. MIT Open Source Building Alliance — A house_n Initiative, Position Paper, MIT House_n, 2003.

[8] Linner T., Güttler J., Bock T. and Georgoulas, C. Assistive robotic micro-rooms for independent living. Automation in Construction, 51(2015):8-22, 2015.

[9] Ojala, T., Pietikäinen, M. and Mäenpää, T. Multiresolution gray-scale and rotation invariant texture with local binary patterns. IEEE Trans. Pattern Analysis and Machine Intelligence, 24(7):971-987, 2002.

[10] Ahonen T., Hadid A. and Pietikäinen, M. Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Analysis and Machine Intelligence, 28(12):2037-2041, 2006.

[11] Aoyama S.K. and Aoki T. A study of biometric recognition Algorithm based on local phase features. In Proceedings of Biometrics Workshop, The Institute of Electronics, Information and Communication Engineers, pages 92-98, 2012.

[12] Terashima H. and Kida T., “勾配情報を用いたLocal Binary Patternの改良”, DEIM Forum 2014, F5-4, 2014.AuthorSurname A. and AuthorSurname B. A journal article. *The Journal*, 1(2):123–345, 2014.