Multi-action Detection System Using Infrared Omnidirectional Cameras

Takashi Imabuchi
Iwate Monodukuri and Software Integration Technology Center (i-MOS), Office of Research and Regional Cooperation, Iwate Prefectural University
Takizawa, Japan
e-mail: t_ima@ipu-office.iwate-pu.ac.jp

Yoshitoshi Murata, Oky Dicky Ardiansyah Prima
Faculty of Software and Information Science
Iwate Prefectural University
Takizawa, Japan
e-mail: {y-murata, prima}@iwate-pu.ac.jp

***Abstract*—Japan is facing a shortage of healthcare workers due to the declining birthrate and aging population. This issue is placing a heavy burden on them to address the growing demand for 24-hour medical services. Information and Communication Technology (ICT) has opened up the possibility of collecting valuable data and providing insights for more control over patients' lives. Early approaches to detect accidental falling and wandering behavior using ICT include the use of invasive and non-invasive sensors. However, in order to put these approaches into practical use, further measures are needed. In this study, we propose a camera monitoring system to automatically detect typical patients' behaviors, such as, rising from the bed, leaving the bed, falling down, and wandering. Our system utilizes infrared omnidirectional cameras that allows a wide range of monitoring actions during day and night. Skeletal information of multiple patients is captured using a computer vision-based pose detection to classify each behavior. Evaluation experiments demonstrated the feasibility of detecting typical patients' behavior using the proposed system.**

Keywords-hospital; patient monitoring; healthcare facility; ommnidirectional camera; human pose estimation.

I. INTRODUCTION

The increasing percentage of elderly people in many national populations [1] is resulting in an increasing number of functionally impaired hospitalized patients, such as, cerebrovascular patients who are paralyzed on one side. Such patients have an increased risk of falling and consequently injuring themselves [2]. Falling is one of the main reasons for them to be hospitalized or placed in residential care. There is also an increasing number of dementia patients, who have a tendency to wander. Inpatient falling and wandering are serious problems for the management of healthcare facilities. Nursing patrols and nighttime monitoring as countermeasures can interfere with patient sleep. The increasing number of patients and the growing shortage of healthcare workers could lead to the stopping of such services. This would result in a lower quality of patient life.

Several measures have been introduced in healthcare settings to detect patient falling and wandering, but their effectiveness is limited. For example, a pressure-sensitive mat sensor placed on or next to a bed can detect the patient leaving the bed but cannot detect wandering or falling down. Moreover, a patient can easily remove or move the mat to prevent sensing, and frequent false detections can results from visitors stepping on the mat. Martinez et al. developed a monitoring system for patients on the bed in a healthcare facility using an infrared camera [3]. It can only detect a patient leaving the bed; it cannot detect wandering or falling. Furthermore, the monitoring area is limited to a bed and its immediate area.

At the research level, Murata et al. developed a multi-action monitoring system for healthcare facilities that uses MS-KINECT sensors [4]. However, they cannot detect a person lying on a bed because they cannot detect differences in the depth between a patient and a bed. Moreover, their coverage is limited to a bed and the surrounding area.

We have developed a monitoring system that uses an infrared omnidirectional camera to automatically detect typical actions by multiple patients in healthcare facilities. Using an infrared omnidirectional camera enables it to monitor patients throughout day and night. It continuously detects and tracks patients on the basis of their skeletal information, estimates the kinds of actions, and notifies the staff if it detects a predefined abnormal action, such as falling.

This paper is organized as follows. Section II describes related work on monitoring systems for the healthcare sector. Section III introduces our proposed system. Sections IV and V describe the experimental setting and results. Section IV discusses the accuracy of estimating a patient's location from an omnidirectional camera image, and Section V describes the evaluation of activity classification based on a patient's skeletal information. Finally, Section VI summarizes the key points and mentions future work.

II. RELATED WORK

Several types of monitoring devices using various kinds of sensors have been introduced in the healthcare sector to detect such patient actions as falling and wandering, as illustrated in Figure 1 [5].

Pressure-sensitive bedside sensor mats (Figure 1 (a)) are commonly used for detecting a patient rising from and/or leaving the bed [6]. Changes in sensor voltage are used to detect rising from the bed, leaving the bed, and standing beside the bed. However, pressure-sensitive mats typically have low durability, can produce false detections due to visitors stepping on them, and can miss detections due to unintentional or intentional movement of the mat.

Clip-type sensors (Figure 1 (b)) have long been used to detect patient leaving the bed. One end of a cord is clipped to the patient's clothing, and the other end is attached to the bed

 ISBN: 978-1-61208-763-4

frame with a magnet. If the patient leaves the bed, the cord detaches from the bed, and a notification is sent to the nurse's station. However, they can cause the patient to feel like they are being monitored, can produce false detections due to patient movement in the bed, and can only detect the patient leaving the bed, not falling.

Heat and infrared sensors (Figure 1 (c)) can detect the patient's location and patient wandering, but they can also detect visitors.

Electromagnetic tags (Figure 1 (d)) are useful for tracking patients [7][8], but indoor positioning accuracy is generally poor due to radio interference. Moreover, patients sometimes refuse to wear a tag or deliberately remove them.

A promising alternative to these methods is non-invasive monitoring using optical sensors, such as, web cameras, because patients cannot disable them. Such sensors are well suited for healthcare facilities because they can capture images of multiple patients simultaneously. However, most such systems currently in use do not support 24-hour monitoring, only daytime monitoring. Moreover, their coverage is limited to a bed and the surrounding area.

Depth cameras are commonly used to estimate the human pose in three dimensions, and the depth camera in an MS-Kinect device has shown adequate performance in healthcare imaging applications [9][10]. They can measure not only the changes in body posture, but also pose (skeletal) information for the targeted person. It is possible to estimate a patient's action using this information.

Recent computer vision applications enable the detection of 2D human poses from a single image [11][12]. Furthermore, the 3D human pose can be estimated by using human pose libraries taken from motion capture devices as a reference [13]. Unlike an MS-Kinect sensor, which must be within a certain distance to the target, these applications are more flexible. Skeletal information can be derived for a body located more than 5 m from the camera. Their application in various fields is expected.

III. Patient Monitoring System

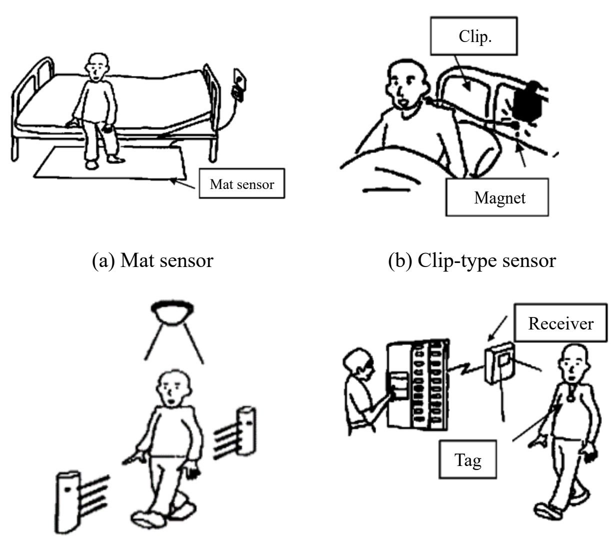
We considered the following requirements to be essential for a patient monitoring system.

- Monitoring both day and night (i.e., 24-hour monitoring).
- Locating and identifying multiple patients.
- Monitoring entire multi-patient room using a minimum amount of easy-to-install equipment.
- Detecting multiple actions, including rising from the bed, leaving the bed, falling down, and wandering.
- Notifying hospital staff of abnormal patient behaviors.
- Protecting privacy.

On the basis of these requirements, we developed a novel patient monitoring system for healthcare facilities. The proposed system detects patient actions using a single infrared omnidirectional camera positioned to face the patient's bed, as illustrated in Figure 2.

An infrared camera is used, which enables 24-hour monitoring. In addition, employing an infrared camera leads to privacy protection because it does not capture clear images compared to a visible-light camera. It has an omnidirectional lens, which enables simultaneously detection of multiple patients in a large patient room. The system continuously and simultaneously detects and tracks multiple patients on the basis of their 2D skeletal information estimated from the camera image. Each patient's actions are classified and labeled in accordance with predefined rules. When an abnormal action is detected, a notification is sent to a hospital staff.

We intend to implement this system on a single-board computer in the near future and complete each image processing as the edge. This will enable easy equipment installation and privacy protection.



(a) Mat sensor (b) Clip-type sensor

(c) Heat and infrared sensors (d) Tag sensor

Figure 1. Various types of sensors used for detecting wandering.

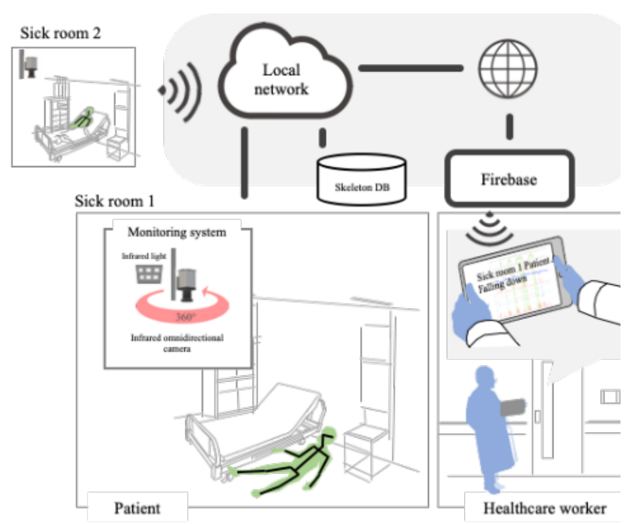


Figure 2. Overview of proposed patient monitoring system.

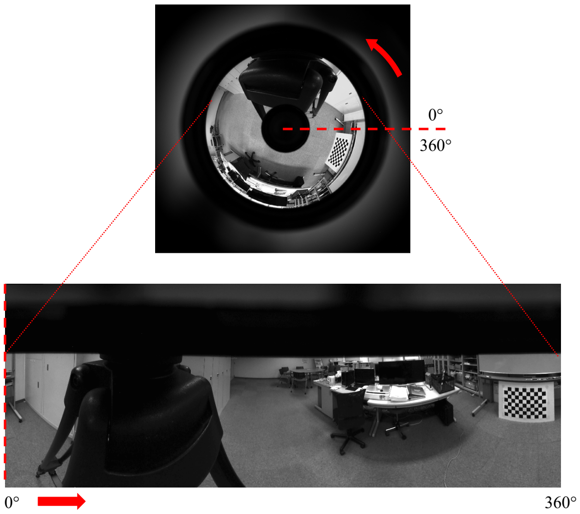


Figure 3. Example expansion of omnidirectional camera image.

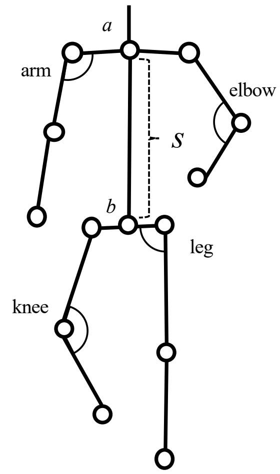


Figure 4. Definition of parameters from skeletal information.

In this paper, we describe the monitoring of patients using an omnidirectional camera. In particular, we describe in detail panorama expansion from images captured by the camera, the detection and tracking of patients using 2D skeletal images obtained using the OpenPose library, joint angle calculation, and action classification.

A. Panorama Expansion

The use of a camera with an omnidirectional lens composed of a hyperboloid mirror enables the capture of a 360° image from the projection of the hyperboloid mirror. With this type of lens, image resolution is high on the sides of and below the lens; the area immediately above the lens cannot be captured. Each captured image is expanded to a panorama image by perspective projection transformation. This requires equal division in the circumferential direction from a predefined center point (x_c, y_c) in the omnidirectional image. Four vertex pairs are calculated as parameters using

$$x = -\frac{a^2 f X}{(b^2+c^2)Z - 2bc\sqrt{X^2+Y^2+Z^2}} + x_c\ , \tag{1}$$

$$y = -\frac{a^2 f Y}{(b^2+c^2)Z - 2bc\sqrt{X^2+Y^2+Z^2}} + y_c\ , \tag{2}$$

where X, Y, and Z are points in 3D coordinates in the omnidirectional image, x and y are points in the image coordinate system on the panorama image, a, b, and c are parameters for the hyperboloid mirror satisfying $c^2 = a^2 + b^2$, and f is the focal length of the camera. These coordinate pairs and perspective projection transformation are used to calculate the 2D coordinate points in the panorama image. Figure 3 shows an example omnidirectional camera image and the expanded panoramic image.

B. Detecting and Tracking Patient Location

The relative location and orientation of the patient's body from the camera are estimated using the 2D skeletal image for the patient. The location of the patient is defined as a polar coordinate P (θ,R), which is calculated using skeletal coordinate *Joint* ($j_1 \ldots j_n$). Figure 4 shows the definitions of the skeletal parameters used in this study. Azimuth θ is centered at the camera and is determined by the ratio indicated by the horizontal coordinate of the Body Center Of Mass (BCOM) under the assumption that horizontal width W of the panoramic image is 360° in all directions. Thus,

$$\theta = 360 \cdot \frac{BCOM_x}{W}\ . \tag{3}$$

The horizontal coordinate of BCOM is calculated using

$$BCOM_x = \frac{j_{1x}+j_{2x}+j_{3x}\cdots j_{nx}}{n}\ . \tag{4}$$

Distance R is determined by the scale of body torso s calculated from the distance between joint a and joint b, as shown in Figure 4. Under the assumption that the posture of a person's body is always parallel to the vertical axis of the camera, R can be determined by multiplying s by the weight calculated on the basis of the ratio between and distortion measured using a calibration process.

C. Calculation of Joint Angles and Action Classification

Patient actions are classified using several parameters calculated from the skeletal information: joint coordinates of patient's head, body variation (standard deviation of BCOM), body axis tilt angle, and joint angles from shoulder, elbow, knee, and leg. First, the standard deviation of BCOM indicating the patient's movement is calculated using

$$Var_{bcom} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(BCOM_{j(x,y)} - \overline{BCOM}_{(x,y)}\right)^2}\ . \tag{6}$$

Then, the tilt angle of the body axis is calculated from the body torso s vector described in Section III-B and the horizontal axis vector.

 ISBN: 978-1-61208-763-4

TABLE I. ACTION CLASSIFICATION RULES

	Inside the bed	Moving	Tilt angle [°]	Joint angles [°]	head position
Rising from the bed	Yes	No	<±30	-	high → low
Leaving the bed	Yes → No	Yes	±30<= → <±30	leg<45 → 150<=leg	-
Falling down	No	Yes → No	±45<=	-	-
Wandering	No	Yes	<±30	160<kne, 150<=elbow	-

The joint angle is calculated as the relative angle between the longitudinal axis of two adjacent segments. For the elbow joint angle, the adjacent segments are the upper arm and forearm. For the arm (shoulder) joint, the adjacent segments are the upper arm and shoulder. For the knee joint angle, the adjacent segments are the upper leg and lower leg. For the leg (hip) joint, the adjacent segments are the upper leg and hip. Let u and v be vectors representing two adjacent segments. Each joint angle between u and v is calculated using

$$\theta_{joint} = 180° - \frac{\bar{u} \cdot \bar{v}}{|\bar{u}| \cdot |\bar{v}|} . \tag{5}$$

Each joint angle is calculated separately for the left and right sides.

The region of the bed is defined such that the bed is arranged with the long side perpendicular to the camera vertical axis.

Using the parameters described above, we set the rules for the four actions considered in this study (see Table I) and classify the actions on the basis of these rules. The results of action detection are labeled and stored as time series data.

IV. EXPERIMENT I

Experiment I was conducted to measure the accuracy of patient location estimation. Using an omnidirectional camera installed at a height of 2.7 m in a multi-patient room, we captured images spanning an arc of 90°. We measured the position at a total of 35 points in increments of 15° up to 90°, 1 to 5 m in increments of 1 m. The camera was an industrial camera (TXG-50, Baumer [14]) with a resolution of 2840 × 2040 and a speed of 30 fps. It was equipped with a PALNON panoramic lens (elevation 66°, depression 0°). Figure 5 shows an example of the panoramic image captured by an omnidirectional camera and a 2D skeleton coordinates of a participant in the experiment. In Figure 5, the left half of the image is hidden by a tripod.

12 participants participated in the experiment. Figure 6 shows the coordinates of reference points (Black cross marker) and the corresponding points of position estimation (Blue circle marker). In addition, red cross markers show averaged for each reference point. For each participant, we calculated the azimuth and distance using the proposed system. The errors were calculated as the Root Mean Square Error (RMSE),

$$RMSE_{\theta} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \sqrt{\left(\hat{\theta}_j - \theta_{i,j}\right)^2} , \tag{6}$$

Figure 5. Example panorama image.

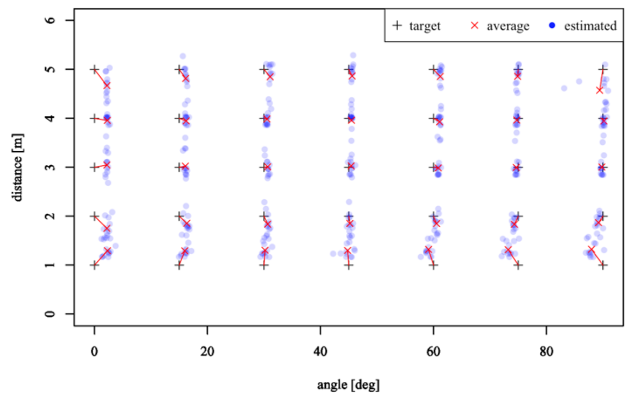


Figure 6. Results of position estimation.

where n is the number of participants and m is the number of reference points. The RMSE for R was calculated in the same way. The RMSE for the azimuth was 1.30°, and the average error for the distance was 0.27 m. These errors are sufficiently small for the proposed system to be used in practical situations because the accuracy is sufficient for estimating the locations of patients between beds in a multi-patient room given that the proposed system enable multiple patients to be simultaneously detected using a single omnidirectional camera. In addition, as shown in Figure 6, no significant error was observed for up to 5 m, indicating that a single omnidirectional camera can effectively cover a large multi-patient room.

V. EXPERIMENT II

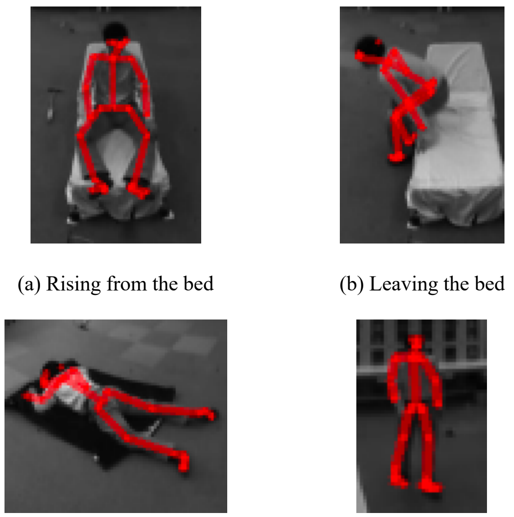
Experiment II was conducted to evaluate the accuracy of the proposed action classification method. Patient actions were collected as video data by having the participants perform the four target actions (rising from the bed, leaving the bed, falling down, and wandering), three times each for each participant. For each video frame, we classified the action using the participant's skeletal information on the basis of the rules given in Table I. The results were calculated as a confusion matrix showing how well the actions were correctly estimated for all frames in the video. We calculated the accuracy and precision from the confusion matrix. The correct data was manually annotated while the video was being checked.

Table II shows the results of action classification as a confusion matrix for eight participants. Figure 7 shows an example of each action captured by the omnidirectional camera. The accuracy of estimating each action exceeded 80%,

TABLE II. RESULTS OF ACTION CLASSIFICATION AS CONFUSION MATRIX

[frame]

	Confusion matrix					
	True positive	False positive	False negative	True negative	Accuracy	Precision
Rising from the bed	601	3366	959	22511	84.24%	15.15%
Leaving from the bed	1087	539	828	24983	95.02%	66.85%
Falling down	2216	961	449	23811	94.86%	69.75%
Wandering	3727	2012	2231	19467	84.54%	64.94%



(a) Rising from the bed (b) Leaving the bed

(c) Falling down (d) Wandering

Figure 7. Example actions and skeletal information.

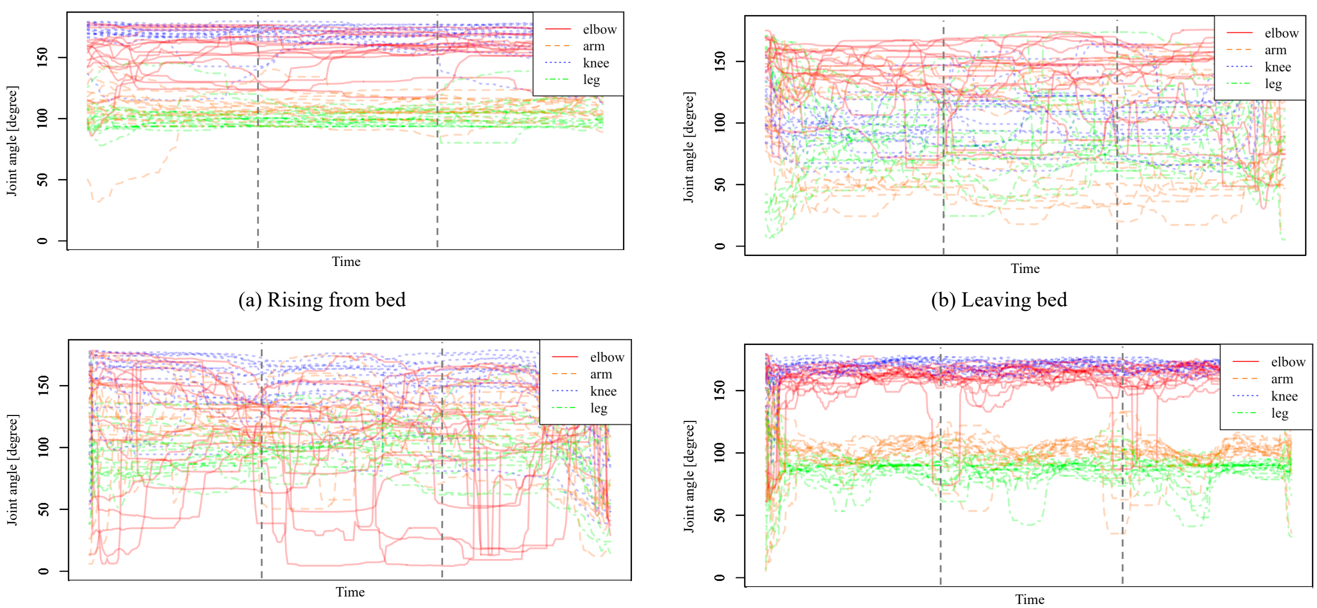
and the precision was 60–70% except for rising from the bed. The number of false detections was high for each action, and the rising from the bed action could not be estimated for most frames. One reason for this is that the head movement in the vertical direction was not large. It is difficult to measure depth information for the joint coordinate occulted by other body parts; and difficult to estimate the body orientation. We found that our rule-based action classification using 2D skeletal images has limitations. We plan to develop a stereo type omnidirectional camera to get 3D skeleton images.

The change in the joint angle for each action was quite similar for all participants. Figure 8 shows graphs of the overlaid changes in joint angle (elbow, arm, knee, and leg; both sides) with the time scale aligned across participants for the four actions.

VI. CONCLUSION

Our proposed patient monitoring system using an infrared omnidirectional camera for healthcare facilities enables detection and classification of various actions that can be dangerous for patients, such as, falling and wandering. Experimental results demonstrated that this system can accurately estimate the locations of multiple patients, enabling each patient to be identified in a wide area. This system should be applicable not only to healthcare facilities but also to facilities that have wide areas such as, factories and schools for use in detecting dangerous situations.

Given the weakness of action classification, we plan to investigate the effect on classification accuracy of the use of machine learning to estimate the changes in joint angles. In addition, we plan to investigate the effect of using a stereo camera to obtain 3D images. For practical application, we will continue to work on improving the accuracy of action



(a) Rising from bed (b) Leaving bed

(c) Falling down (d) Wandering

Figure 8. Overlaid change in joint angle with time scale aligned across participants.

classification, using only joint information for personal authentication to ensure anonymity, and integrating the system with our developed alert notification systems.

REFERENCES

[1] World Population Ageing: 1950-2050, Population Division, Department of Economic and Social Affairs, United Nations, http://www.un.org/esa/population/publications/worldageing19502050/ [retrieved: Febrary, 2020].

[2] W. P. Berg, H. M. Alessio, E. M. Mills, and C. Tong, “Circumstances and consequences of falls in independent community-dwelling older adults,” Age Ageing, Vol. 26, pp. 261–268, 1997.

[3] M. Martinez and R. Stiefelhagen, “Automated multi-Camera system for long term behavioral monitoring in intensive care units,” 2013 IAPR International Conference on Machine Vision Applications, pp. 97–100, 2013.

[4] Y. Murata, R. Takahashi, T. Yamato, S. Yoshida, and M. Okamura, “Novel Field Oriented Patient Monitoring Platform for Healthcare Facilities,” IARIA, International Journal on Advances in Software, Vol. 11 No. 3&4, pp. 368–378, 2018.

[5] I. Kitayama, K. Omori, H. Matsuno, and Y. Sugimoto, “Wandering prevision and monitoring systems for persons with dementia (Part 2) - Usage survey of wandering prevision and monitoring system/machine at nursing homes and other welfare facilities-,” Report of the Hyogo Prefectural Town Welfare Research Institute, pp. 98–111, 2003.

[6] Mat sensor, Carecom, https://www.carecom.jp/global/solutions/option/ [retrieved: Febrary, 2020].

[7] T. Gao, D. Greenspan, M. Welsh, R. Juang, and A. Alm, “Vital signs monitoring and patient tracking over a wireless network,“ Johns Hopkins APL Techincal Digest, Vol. 27, No. 1, pp. 66–74, 2006.

[8] P.Varady, Z. Benyo, and B Benyo, “An open archtectre patient monitoring system using standard technologies,” IEEE Tracnsaction on Information Technology in Biomedicine, Vol. 6, No. 1, pp. 95–98, 2002.

[9] E. Gambi, et al., “Heart rate detection using Microsoft Kinect: Validation and comparison to wearable devices,” Sensors 2017, Vol. 17, No. 8, pp. 1776, 2017.

[10] S. T. L. Pohlmann, E. F. Harkness, C. J. Taylor, and S. M. Astley, “Evaluation of Kinect 3D sensor for healthcare imaging,” Journal of Medical and Biological Engineering, Vol. 36, pp. 857–870, 2016.

[11] A. Newell, K. Yang, and J. Deng, “Stacked Hourglass Networks for Human Pose Estimation,” European Conference for Computer Vision (ECCV) , pp .483–499, 2016.

[12] Z. Cao, T. Simon, S.E. Wei, and Y. Sheikh, “Realtime multi-person 2D pose estimation using part affinity fields,” Computer Vision andPattern Recognition 2017, pp. 1302–1310, 2017.

[13] H. J. Lee and Z. Chen, “Determination of 3D human body postures from a single view,” Computer Vision, Graphics and Image Processing, Vol. 30, pp. 148–168, 1985.

[14] Baummer, https://www.baumer.com/ch/en/ [retrieved: Febrary, 2020]