Privacy-Preserving Monitoring System with Ultra Low-Resolution Infrared Sensor

Miyuki Ogata[1], Shogo Murakami[1], Kimura Takumi[1], Ikuko Eguchi Yairi[1]
[1]Department of Information and Communication Sciences, Sophia University, Tokyo, Japan
{ogata | murakami | kimurat}@yairilab.net, i.e.yairi@sophia.ac.jp

CEUR-WS.org/Vol-2760/paper5.pdf

Abstract

Action monitoring systems used in households provides vital information for health monitoring particularly with aging residents. While visual inputs such as information provided by cameras can recognize the actions and position of a subject with high accuracy, they are not widely accepted due to privacy concerns. This paper proposes a posture classification method with the use of a low-resolution thermal sensor. The sensor aims to protect the subject's privacy by capturing visual input in the infrared spectrum as well as having a low spatial resolution of 8x8 pixels. We consider a simulation which recreates the experimental environment and produces data for this postural-behavioral problem. The validity of this method is checked by considering 3 postures; standing, sitting, and laying down and examined using a classifier on simulated data. Additionally, we explore optimal position and angle of the sensor as well as the effects of color depth on accuracy. In our results we achieve over 93% classification accuracy by color conversion of the infrared array sensor image and successfully decreased loss due to displacement by DCNN. We discover higher accuracies are achieved when the sensor is located 50cm below the subject's height with a tilt angle of ±2°.

1 Introduction

Japan, like many other countries, is said to have a rapidly aging society with more than 26.6% of its population being over the age of 65 [Statistics Bureau, 2019]. As an aging society increases, so does the incidence of suffering from medical conditions which includes but is not limited to; cardiovascular, musculoskeletal, and cognitive disorders, as well as other chronic diseases. These diseases often require around the clock supervision traditionally provided by caregivers or family members. In addition to this, Japan has the issue of "kodokushi" or "lonely deaths" which refers to the phenomenon of people dying alone and often remaining undiscovered for long stretches of time. This phenomenon has garnered attention of the public as the known incidents of lonely deaths continue to grow [Cabinet Office, 2016]. Full time supervision and assistance of the elderly population is needed to prevent these kinds of deaths however; it often comes at a high price. Care facilities have the advantage of offering 24-hour care, but the economic burden becomes higher the longer the patient remains. Conversely, it is impossible for a single caregiver to tend to a senior resident at all times thus being inefficient in preventing sudden incidents. For these reasons, demand for a reliable monitoring system to help improve health care has grown.

Supervision with a monitoring system through the use of IoT would present a more adequate and cheap solution to human alternative. Once installed, IoT devices can be used at all times in order to ensure the safety of the user. The two methods that are currently available are wearable sensors and visual sensors.

Improvements in artificial intelligence and the spread of the internet has made behavior analysis more accessible. Wearable sensors focus on obtaining behavior analysis at a low cost, low energy consumption and provides data simplicity [Mukhopadhyay, 2015]. Devices such as smartphones and smartwatches can detect falls, share its location, or record cardiac beats [Murad et al., 2017; Won-Jae *et al.,* 2014; Najafi *et al.,* 2003]. In general, the tri-axial accelerometers are used to analyze the movements and position of the device wearer. Over recent years there has been a rapid development in this technology, yet issues surrounding this technique still remain. These include limited battery life, poor comfort, extended period of wear, and irregular wearing. The latter is the most problematic in the elderly population as many suffer from illnesses such as dementia, Alzheimer or simply becoming forgetful and as such, overlooking to wear these devices.

Sensors can also be embedded into daily items for monitoring and activity recognition purposes. A particular object's use frequency can be recorded to detect abnormal activity. For example, a sensor in a cane would register an abnormal movement if the cane were to suddenly fall to the ground [Vahdatpour *et al.,* 2010]. While this method protects the user's privacy, monitoring is limited to scenarios in which

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

these items are being handled. Multiple devices can be tracked to increase broadness, leading to increased complexity and high costs.

Visual sensors can capture high resolution images of the user's daily activity. This method contains information of multiple events concurrently, which allows for a rich data behavior analysis. This in turn outputs a very accurate recognition of behaviors. In an ever-growing digital age, being recorded constantly can be a deterrent for many users in the form of privacy concerns. To eliminate behavioral limitations and physical stresses for the user, infrared sensors can be employed in monitoring systems. These sensors only detect temperature information, making its use and installation simple while keeping privacy breaches to a minimum [Okada and Yairi, 2013].

Infrared (IR) monitoring devices are categorized into those using single beam sensors and those using an array of sensors. While single beam IR sensors sense temperature at a single point, IR array sensors are comprised of multiple single beams to obtain deeper spatial information. IR array sensors obtain spatial information in three dimensions and receive light magnitude within a specified region to make posture estimations of users [Hayashida *et al.*, 2017]. Current concerns with IR sensors include the accuracy and sensitivity of array sensors to changes in background temperature, installation position, and installation angle. As of today, studies have only recorded data in optimal conditions. Consequently, the impact of variances in position and angle of IR sensor is not known.

This paper focuses on the classification accuracy of three postures with the use of DCNN and a single 8x8 infrared array sensor. Experimental and simulated data is used to investigate its effectiveness as a privacy preserving monitoring system.

2 Related Works

The use of computer vision to assist in personal well-being and reduce incidents at home has increased in demand over the last decade. Studies focused on wearable devices and high-resolution monitoring systems have been widely explored. In contrast, reports on privacy preserving vision-based monitoring systems are still quite limited. Low-resolution thermal sensors can detect temperature with the use of a small array of infrared sensors. This produces a spatial distribution of temperature which can be represented in the form of low-resolution images. While cameras are able to distinguish an individual's features, infrared sensors can only capture the outline or shape of the human body, making identification of an individual difficult. The use of low spatial resolution further decreases the image resolution making identification nearly impossible. Utilization of this sensor is the most suitable for homes as it provides a more comfortable experience by being small, unobtrusive, and cheap while protecting an individual's privacy.

Previous studies for detection, counting, and tracking of people have been performed with an 16x16 infrared sensor for indoor monitoring [Berger and Armitage, 2010]. Recognition of hand motion direction has also been investigated using a 4x4 infrared sensor [Wojtczuk *et al.,* 2011] however this extremely low resolution would not be suitable for more complex visual recognition.

Activity monitoring using an 8x8 infrared sensor has been successfully researched before [Tao *et al.,* 2018]. However, these studies have recorded data in optimal conditions only. External factors are often not considered in the sensor and systems ability to perform. Effectiveness of these devices could be limited by both the location of installation and installation angle. The contribution of color depth in the accuracy of classification is also not thoroughly studied in the field of monitoring systems. Additionally, we need to consider the difficulty regarding the acquisition of learning data to increase classification accuracy. The investigation of participants under different conditions such as room size, room temperature, body type, and posture would require extensive preparation and could lead to inaccuracies. Collecting data from actual participants in this way would be unrealistic due to time limitations and the high costs involved.

3 Proposed Method

3.1 Device Setup

The device shown in Figure 1 is used for data collection of posture classification. This device is comprised of an infrared array sensor Grid-EYE by Panasonic mounted on a series of single-board computer, Raspberry Pi3 Model B. The Grid-EYE sensor has an output of 8x8 data for surface temperatures detected in the observed space. This data can be visually represented as an 8x8 greyscale image. The sensor detects temperatures from 0°C to 80°C with a temperature resolution of 0.25°C. Its viewing angle is 60° with a sampling rate of 10fps [Panasonic, 2016].

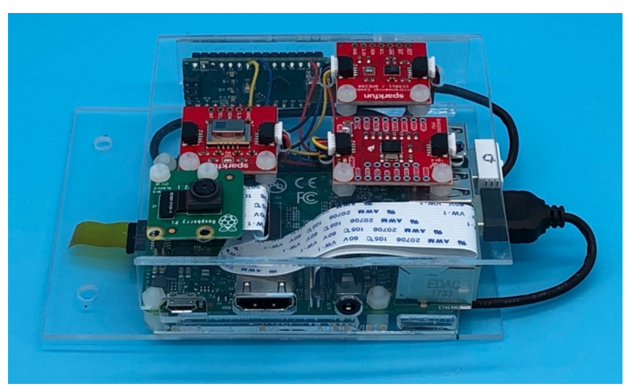

Figure 1: Action recognition device with 8x8 infrared sensor

3.2 Deep Convolutional Neural Network (DCNN)

An effective method for obtaining high accuracy in image recognition is with the use of Deep Convolutional Neural Network (DCNN). In this study we implement DCNN to extract features from infrared image data in order to train the posture classifier. The network contains five layers: an input layer, two convolutional layers, a fully connected layer, and an output layer [Kimura *et al.,* 2019]. Max pooling was implemented in the convolution layers [Yang *et al.,*2015]. Dropout was used to prevent over learning and Adam optimizer was employed to update parameter weighting to optimum levels [Kingma and Ba*,* 2014].

3.3 Experimental Setting and Experimental Dataset

Our experiment was conducted on 3 subjects to evaluate the performance of DCNN on posture classification. The position of two subjects; one male (age 24) with a height of 170cm and one female (age 20) with a height of 160cm were recorded in a 9.5m^2 Japanese-style room with ambient temperature of 13°C. The position of the third subject; one male (age 22) with a height of 170cm was recorded in a 20m^2 room with ambient temperature of 11°C. In all three set-ups, the sensor was placed 140cm form the ground so that the subjects' entire body could be observed. Subjects were told to stand, sit, or lie down remaining within 1m to 3m from the sensor so as to remain within the sensor viewing range. A total of 14,983 frames worth of data was recorded, a randomized 10% of which was used to train the classifier and the remaining 90% used for testing. Accuracy was found to plateau at around 100 epochs, consequently it was set as such. Since the sensor does not discriminate between subjects, posture classification result and accuracy evaluation are the combined dataset from the three subjects. The results are as shown in Table 1.

For real-world use, resident monitoring systems should be able to detect sudden incidents such as slips or falls, therefore the F-measure should be at or above 90%. Experimental results showed an average F-measure of 87%. By observing the classification results we found that incorrect categorizations were predominantly in postures going from standing to seated, and from seated to laying down.

3.4 Color conversion on Classification Accuracy

The values from the infrared sensor were directly inputted into the DCNN. However, due to the quantity of erroneously categorized data, it can be hypothesized that without further processing, an accurate feature extraction cannot be achieved. A previous study by Ito [2018], reported on on-board camera and deep learning for pedestrian crossing detection. In it, they converted greyscale data into colored images with the addition of edge processing and were able to significantly increase accuracy [Ito *et al.,*2018]. Similarly, our DCNN is able to handle multiple channel inputs, including color images. The 8x8 infrared sensor values can be converted into 8x8 colored images, which will then be pre-processed to input into the DCNN. The conversion into different color spaces such as RGB, HSV, CIE XYZ, and CIE Lab can be used to increase feature recognition [Rachmadi and Purnama, 2015]. Other studies have also used a combination of color spaces to create an efficient face recognition system [Kurylyak *et al.,* 2009]. Since the mentioned studies only use color conversion on high resolution image data, its effects on low resolution image data are explored in this study. The most effective color conversion for this subject will be chosen through the processing of data into various color spaces and comparing results.

Posture classification results

Recog-nition	Answer: Stand	Sit	Lie
Stand	3582	266	15
Sit	496	5243	435
Lie	185	590	4171

Posture classification evaluations

	Stand	Sit	Lie
Accuracy	0.9272	0.8492	0.8433
Recall ratio	0.8402	0.8596	0.9026
F-measure	0.8816	0.8544	0.8719

Table 1: Classification result and evaluation from infrared sensor images by DCNN

Posture classification results

Recog-nition	Answer: Stand	Sit	Lie
Stand	3929	137	20
Sit	217	5620	113
Lie	121	336	4490

Posture classification evaluations

	Stand	Sit	Lie
Accuracy	0.9615	0.9445	0.9076
Recall ratio	0.9207	0.9223	0.9712
F-measure	0.9407	0.9333	0.9383

Table 2: Classification evaluation from 8x8 RGB converted images.

The 8x8 greyscale data image were converted into color to evaluate classification accuracy. Color spaces as determined by the CIE (Commision Interon acznationale de l'Eclairage) such as RGB, CIE XYZ, CIE L*a*b* were assessed. For the RGB color space, the greyscale image is divided into three image layers each representing R, G, B which is fed into the DCNN. The input layer to the DCNN is converted from 1 channel into 3 channels. Temperature data was mapped such that lower temperature values were shown in blues and higher values were shown in reds. This method was repeated for both CIE XYZ and CIE L*a*b* with the use of OpenCV (Open Source Computer Vision Library) developed by Willow Garage Inc.

The same test-train ratio as in the previous section was used to train the classifier. The classification and evaluation process were repeated 100 times to assess their accuracy. Classification accuracy of 93% was achieved by XYZ; 91% was achieved by L*a*b*. The best results were attained by RGB color conversion as shown in Table 2.

4 Experimental and Simulated Dataset

It is hypothesized that a simulated dataset can be created without the need of human participants in order to increase the volume of dataset for learning. A simulation that produces computer-generated data was constructed using Unity, a cross-platform game engine with pre-installed libraries for physical functions. Its versatility for creating the required models as well as altering their conditions such as; modeled humans' height, physique, position, sensor tilt, height, and room size was the main reason for which Unity was selected.

4.1 Assessment of Experimental and Simulated Datasets

The Grid-EYE sensor uses 64 elements to detect temperature through the measurement of emitted infrared light from objects. Furthermore, due to dissipation of heat and background noise, it was noted that increased distance from sensor decreases the precision of temperature measured. In order to replicate this, a virtual sensor was added to the simulation using ray tracing from point of observation. The use of ray tracing helps locate the human model and accurately recreates the physical occurrence caused by distance.

Typically, data collected from simulators require highly detailed human models to reproduce a real-life scenario. However, since this study utilizes a low-resolution sensor device a simple 3D model can replicate the experimental setup. The human modeled was a rudimentary model composed of legs, arms, a torso and a head. Temperature distribution for each body part is set separately so as to closely resemble the temperature distribution of a real human. The room size and walls temperature distribution were set to be analogous to the one present in the experimental setup.
The human model is able to be positioned in either standing, seated, or laying down and was placed in random locations between 1m to 3m away from the sensor.

4.2 Training and Results of Datasets

Infrared image data from participants and simulated data in standing position at different distances is shown in Figure 2. The figure is color coded to display areas of high temperature in reds and areas of low temperatures in blues. Due to the low-resolution nature of the sensor both real data and simulated data are nearly indistinguishable from each other. A minor difference can be perceived in the area surrounding the individual in the real data where temperature gradient appears as a result of heat transfer. In the simulated data a clearer temperature split can be appreciated between the human model and background.

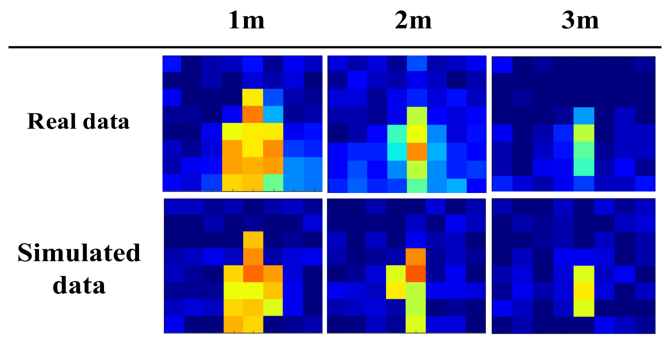

Figure 2: Comparison of sensor data and simulated data for standing subject

In our study, 5000 simulated images were used to train the DCNN posture classifier and experimental data was used to test the classification accuracy. This method achieved a 70-80% classification accuracy which is too low for practical use. Causes for wrong categorization can be due to the heat transfer from the body to surroundings or sources of heat such as light, which increases the temperature of the participants' surrounding area. This background noise is not as prevalent in the simulated data. Elimination of background heat from experimental data should decrease errors. Background data elimination was applied to both sets of data; using the simulated data for training and real data for testing. Results are as shown in Table 4 were F-measure of all categories increased to over 90%.

Posture classification results

Recog-nition	Answer Stand	Sit	Lie
Stand	3306	537	216
Sit	959	5129	1274
Lie	5	427	3130

Posture classification evaluations

	Stand	Sit	Lie
Accuracy	0.8144	0.6966	0.8787
Recall ratio	0.7742	0.8417	0.6774
F-measure	0.7938	0.7623	0.7650

Table 3: Classification result of real data by learning from simulated data

Posture classification results

Recog-nition	Answer: Stand	Sit	Lie
Stand	3929	137	20
Sit	217	5620	113
Lie	121	336	4490

Posture classification evaluations

	Stand	Sit	Lie
Accuracy	0.9615	0.9445	0.9076
Recall ratio	0.9207	0.9223	0.9712
F-measure	0.9407	0.9333	0.9383

Table 4: Classification result of background removed real data by learning from simulated data

4.3 Evaluation of Height and Angle of installation

Thus far, an accuracy of 90% has been achieved by thermal sensor being positioned at 140cm off the ground. For actual applications, the heights of users differ from individual to individual. Additionally, changes in sensor position or tilt can occur due to external causes such as being shaken during an earthquake or being accidentally bumped into. We assessed the sensors' performance with different installation conditions.

Using the simulation, sensor height was varied from heights of 50-170cm in step increments of 10cm. Using the data simulator, the effects were explored on human models of 150cm, 160cm, and 170cm tall. Results shown in Figure 3 revealed classification accuracy on postures reached peak accuracy when sensors were installed at 50cm below the users' height.

The influence of tilt angle of sensor on classification accuracy was also studied. Using a model height of 170cm and sensor set at its optimum 120cm elevation tilt was explored between the ranges of –5° and 5° in incremental steps of 1°. Positive angle describes sensor tilted upwards; negative angle describes sensor tilted downwards. As upwards tilt is increased F-measure tended to decline. The same trend occurs when downwards tilt is applied, however F-measure seems to severely drop after –1°, as seen in Figure 4. F-measure is sustained over the value of 90% between the tilt ranges of –2° and 2°.

Protection of privacy is a serious concern when dealing with monitoring devices. Posture classification used in this research rely heavily on temperature distribution captured by the 8x8 infrared sensor. Due to low resolution of the image, information received in each pixel is of vast importance. Any loss or noise could negatively affect DCNN ability to categorize. Due to tilt, a full row of pixels might be shifted or completely evaded, preventing the full capture of the persons posture. This loss of information will cripple the system's ability to classify postures correctly.

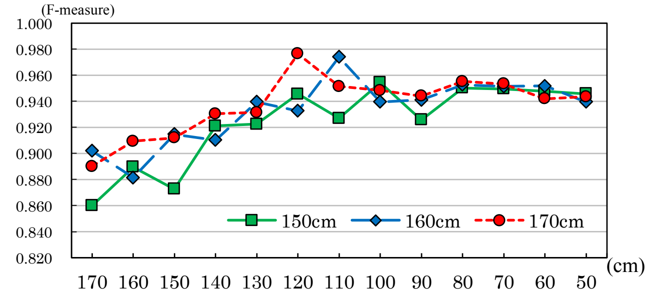

Figure 3: Person height with varying sensor position

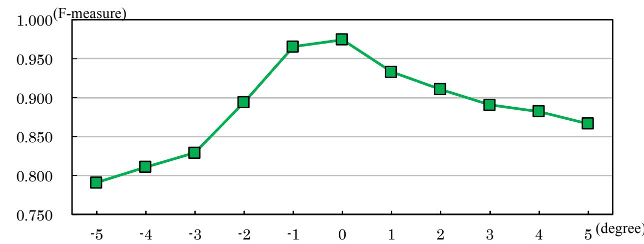

Figure 4: Vertical tilt angle for sensor

5 Discussion

5.1 Household Implementation

For successful commercial use, the supervising system must be easy to install and adaptable to buyers' specification. The device used in this experiment can be attached to an indoor room wall, placed at 50cm below clients' height. If surface is bumpy, blocked or angled it may need further adjustments. Ease of use if another aspect of concern for monitoring devices. In our case, the unobtrusive nature of the infrared sensor device lends to the forgetfulness of seniors, once set and installed there is no need for users to interact with the device.

While our system successfully detects three basic human postures (standing, sitting, and laying down) falls are not included. It is well known that fall action recognition is imperative in health monitoring for elders as quick response is crucial in death prevention. To further differentiate actions such as sitting down versus sitting up and laying down versus falling from three basic postures, we need to consider temporal dimension in our dataset. We propose the addition of temporal feature extraction alongside spatial feature extraction for classification of actions as shown in Figure 5.

5.2 Temporalization of Action Recognition

Integration of Recurrent Neural Networks (RNN) can provide us with the temporal feature maps of input data. RNNs exhibits dynamic temporal behaviors as they can process sequences of input and recognize patterns. The addition of temporalization in our recognition system should provide us with time dependent features extracted from the dataset to recognize actions.

Long Short-Term Memory (LSTM) is an artificial RNN which has been typically used in speech recognition and

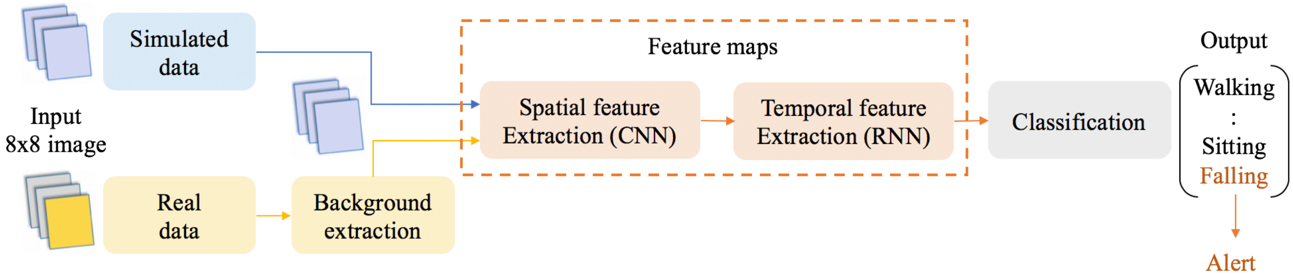

Figure 5: Illustration of proposed human action recognition.

multi-language processing. LSTM in combination with CNNs has been used in automatic image captioning, weather forecast, and emotion recognition. It is our hypothesis that it can also be used for temporal feature extraction in low resolution sensor data.

One of the difficulties in training an RNN for action recognition involves time frames. The time required for an individual to perform movements whether from standing to laying down or standing to falling can vary greatly. This variation depends on factors such as age group, level of mobility, and health status or previous injuries. A study recorded bed rising time (from supine to sitting position) taking an average of 2.5s for adults, 4s for seniors in congregate housing, and 10s for seniors in skilled nursing facilities [Alexander *et al.,* 2000]. Moreover, external factors such as type of fall and trying to stop a fall by holding on to side-rails, canes or other objects can also change the time frame of the fall. Detecting for such inconsistent and irregular falls might be difficult to train for, as they could be outliers in the sample.

5.3 Personalization and Issues

Classification of actions can be divided by age groups, with shorter time frames dedicated to younger/healthier seniors; and longer time frames for older weaker seniors. Alternatively, a broader time frame can be set to encompass all age groups. A drawback for this method is that time frames for other movements might be overlap.

The monitoring system can be further personalized by inputting real data from user detected by thermal sensor device. However, actions such as sitting, standing, and laying down can cause strain when performed repeatedly by senior citizens. Moreover, actions such as falling are potentially dangerous when repeated, making its data collection not feasible. Instead, as shown in Chapter 4.1 the use of data generating simulator can be used to increase dataset for these actions lessening the risk of injury from its users.

The degree to which this system can be personalized is still to be determined. While 8x8 sensor preserves the user information privacy and is faster to process, it also brings forth issues due to its low resolution. Hyper low-resolution sensors inherently have the disadvantage of containing less information compared to their high-resolution counterparts. Low image pixel dimension means that any cropping will lead to significant loss of information.

6 Conclusion

In this paper we explored the performance of infrared array sensor in resident monitoring system. Through the use of 8x8 sensor image we managed to yield over 90% accuracy for human posture classification. We analyzed data noise created by external factors on sensor tilt and position. We concluded that tilt angle within $\pm 2°$ and a position of 50cm below subjects' height returned the highest accuracy. While height variation made F-measure decrease by a maximum of 10%, tilt variation can decrease F-measure by over 25%. This highlights the importance of proper positioning and tilt of sensor according to room size and users' height. These are strong variables and are key to record movement and optimize accuracy.

Additionally, we extended our study by introducing simulated data and found that it is a viable complementary way to increase data sample size. We believe our work is the first to apply a simulation model to increase data for low resolution monitoring in the field of action recognition.

For further research, we can improve the data simulator and learning algorithm by including temporal feature extraction for action recognition and application in real environment. As shown in Figure 5 we hypothesize simulated data can be used to assist on data training without the need of lots of real data recordings. The real data will undergo background removal to broaden our monitoring system capability.

Reference

[Alexander *et al.,* 2000] Neil B. Alexander, Julie C. Grunawalt, Scott Carlos, and Joshua Augustine. Bed mobility task performance in older adults. *Journal of rehabilitation research and development* 37(5):633-638, September 2000.

[Berger and Armitage, 2010] Martin Berger and Alistair Armitage. Room occupancy measurement using low-resolution infrared cameras. *IET Irish Signals and Systems Conference (ISSC 2010),* pages 249-254, June 2010.

[Cabinet Office, 2016] Cabinet Office. Annual Report on the Ageing Society: FY 2019. *Cabinet Office Government of Japan*, June 2019.

[Hayashida *et al.*, 2017] Akira Hayashida, Vasily Moshnyaga, and Koji Hashimoto. The use of thermal ir array sensor for indoor fall detection. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* pages 594-599, December 2017.

[Ito *et al.,*2018] Toshio Ito, Yuki Takata, and Mohd Hafiz Hilman bin Mohammad Sofian. Crossing Pedestrian Detection Using Deep Learning by On-board Camera. *Transactions of Society of Automotive Engineers of Japan,* 49(5), September 2018.

[Kimura *et al.,* 2019] Takumi Kimura, Shogo Murakami, and Ikuko Eguchi Yairi. Privacy-Preserving Resident Monitoring System with Ultra Low-Resolution Imaging and the Examination of Its Ease of Installation. *The Japanese Society for Artificial Intelligence 33rd annual conference*, pages 4D3E201-4D3E201, June 2019.

[Kingma and Ba*,* 2014] D. Kingma and J. Ba. Adam: A Method for Stochastic Optimization, *International Conference on Learning Representations*, Dec 2014.

[Kurylyak *et al.,* 2009] Yuriy Kurylyak, Ihor Paliy, Anatoly Sachenko, Amine Chohra, and Kurosh Madani. Face detection on grayscale and color images using combined cascade of classifiers. *Computing = Комп'ютинг* 8(1), pages 61-71, 2009.

[Mukhopadhyay, 2015] Subhas C. Mukhopadhyay. Wearable sensors for human activity monitoring: A review. *IEEE sensors journal*, 15(3):1321-1330, March 2015.

[Murad et al., 2017] Abdulmajid Murad and Jae-Young Pyun. Deep recurrent neural networks for human activity recognition. *Sensors Journal,* 17(11):2556, November 2017.

[Najafi *et al.,* 2003] Bijan Najafi, Kamiar Aminian, Anisoara Paraschiv-Ionescu, François Loew, Christophe J. Bula, and Philippe Robert. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *IEEE Transactions on biomedical Engineering* 50(6):711-723, June 2003.

[Okada and Yairi, 2013] Ryotaro Okada and Ikuko Yairi. An indoor human behavior gathering system toward future support for visually impaired people. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, article 36, October 2013.

[Panasonic, 2016] Panasonic. Infrared Array Sensor Grid-EYE specification sheet. http://www.farnell.com/datasheets/2058257.pdf?_ga=2.187765294.767814663.1555052764-1882353189.1555052764 created, May 2016. Last visited Jan 2020.

[Rachmadi and Purnama, 2015] Reza Fuad Rachmadi and I. Ketut Eddy Purnama. Vehicle color recognition using convolutional neural network. *Computing Research Repository Journal (CoRR)*, October 2015.

[Statistics Bureau, 2019] Statistics Bureau. Statistical Handbook of Japan 2019. *Ministry of internal Affairs and Communications*, pages 8-16, May 2019.

[Tao *et al.,* 2018] Lili Tao, Timothy Volonakis, Bo Tan, Yanguo Jing, Kevin Chetty, and Melvyn Smith. Home Activity Monitoring using Low Resolution Infrared Sensor Array. *Computing Research Repository Journal (CoRR)*, November 2018.

[Vahdatpour *et al.,* 2010] Alireza Vahdatpour and Majid Sarrafzadeh. Unsupervised discovery of abnormal activity occurrences in multi-dimensional time series, with applications in wearable systems. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, pages 641-652. Society for Industrial and Applied Mathematics, May 2010.

[Wojtczuk *et al.,* 2011] Piotr Wojtczuk, Alistair Armitage, T. David Binnie, and Tim Chamberlain. PIR sensor array for hand motion recognition. In *Proceedings of the 2nd International Conference on Sensor Device Technologies and Applications*, pages 99-102, August 2011.

[Won-Jae *et al.,* 2014] Yi Won-Jae, Oishee Sarkar, Sivisa Mathavan, and Jafar Saniie. Wearable sensor data fusion for remote health assessment and fall detection. *IEEE International Conference on Electro/Information Technology*, pages 303-307, June 2014.

[Yang *et al.,*2015] J.B. Yang, M.N. Nguyen, P.P. San, X.L. Li, S. Krishnaswamy. Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence* (IJCAI), pages 3995–4001, Buenos Aires, Argentina, June 2015.