Activity Recognition Using Low Resolution Infrared Array Sensor

Shota Mashiyama
Graduate School of
Science and Technology,
Keio University,
Yokohama 223–8522
Email: mashiyama@ohtsuki.ics.keio.ac.jp

Jihoon Hong
Faculty of
Science and Technology,
Keio University,
Yokohama 223–8522
Email: hong@ohtsuki.ics.keio.ac.jp

Tomoaki Ohtsuki
Department of Information and
Computer Science,
Keio University,
Yokohama 223–8522
Email: ohtsuki@ics.keio.ac.jp

***Abstract*—Now, aging society is a worldwide problem, and the population of people aged over 60 years is growing faster than any other age group. Therefore, monitoring services for elderly people are attracting a great deal of attention. We have proposed a fall detection method using a low resolution infrared array sensor to inform an unexpected falling in our previous work. However, knowing daily fundamental activities of elderly people is also important to prevent future accidents. In this paper, we propose an activity recognition method using a low resolution infrared array sensor. This sensor can detect temperature on a two dimensional area. From the viewpoint of general versatility (available in darkness), cost, size, privacy (low resolution), and availability (commercial off-the-shelf), this sensor is better than other sensing devices like video cameras, Doppler radars, acceleration sensors, and so on. In the proposed method, temperature distribution obtained from the sensor is analyzed and classified into five fundamental states: "No event", "Stopping", "Walking", "Sitting", and "Falling" (emergency situation). As a result of experiments, our proposed method achieved recognition accuracy of 100 %, 94.8 %, 99.9 %, and 78.6 % respectively. In particular, 100 % accuracy of "Falling" recognition was achieved.**

I. Introduction

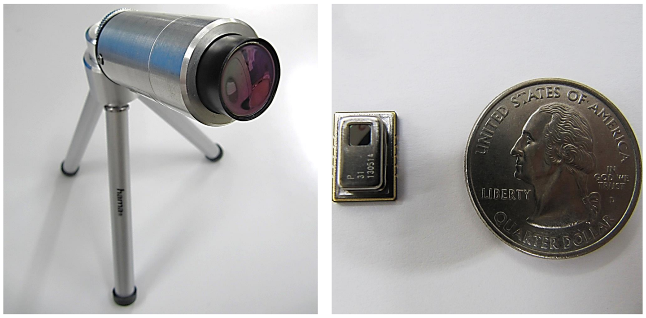
Nowadays, aging society is a big problem in many countries [1]. In particular, the state that more than 21 % of population are occupied by elderly people is called "super aging society." According to Cabinet Office of Japan, the total population of Japan is estimated to decrease in the future. On the other hand, the population aging rate that indicates the proportion of population aged 65 and over is estimated to increase [2]. In this circumstance, the number of elderly people living alone is increasing. Therefore, for family members living away from him/her, there is an increasing number of expectations for elderly monitoring services that can inform their general activities or emergency situation like a fall. Recently, the application of activity recognition advances in various fields including health care, advertisement and video game.

In general, activity recognition can be divided into two types. One is wearable, and the other is non-wearable. Wearable type requires user to have or wear sensing devices. For example, there are wearable activity recognitions using smartphone sensors [3], [4], acceleration sensors [5], [6], capacitive sensors [7], and so on. Advantages to use wearable sensing devices are mobility, and high activity recognition rates due to low noise. However, these methods can be literally used only while wearing, and they are not preferred for elderly people monitoring. In contrast, non-wearable type does not require user to have or wear sensing devices. The devices are mostly fixed in predetermined place. For example, there are non-wearable activity recognitions using array antenna [8], video cameras [9], sensor fusion [10], and binary infrared sensors [11]. Advantages to use non-wearable sensing device are user friendly and service continuity (energy supply and equipment failure rate). However, activity recognition rates of these methods tend to be lower than those of wearable type methods. In addition, the video camera-based method [9] has issues of privacy invasions and unavailability in darkness (unless use a special equipment), and the sensor fusion-based method [10] is complicated and costs a lot. The binary infrared sensor-based method [11] can recognize activities by classifying output patterns obtained from the sensors installed on the ceiling. However, recognition result depends on sensor placement (e.g. When the sensor placed near by PC reacts, the activity is estimated to be "Using PC".), and emergency situation like falling is not detected (there is too little information obtained from the binary infrared sensors).

As the first step of activity recognition, we have proposed a fall detection method using the infrared array sensors in previous work [12]. Advantages of using this sensor are general versatility (available in darkness), low cost, small size, privacy protection (low resolution) and availability. The sensor has two or more infrared detectors inside, and temperature detection is achieved on a two dimensional area. In the previous work, we analyzed temperature distribution, and based on it, classified activities as a fall or a non-fall. However, knowing daily fundamental activities of elderly people is also important to prevent future accidents. Therefore, to overcome aforementioned issues, we propose a novel activity recognition using a low resolution infrared array sensor in this paper. We analyze the temperature distribution obtained from the sensor and classify activities into five classes: no event, stopping, walking, sitting and falling.

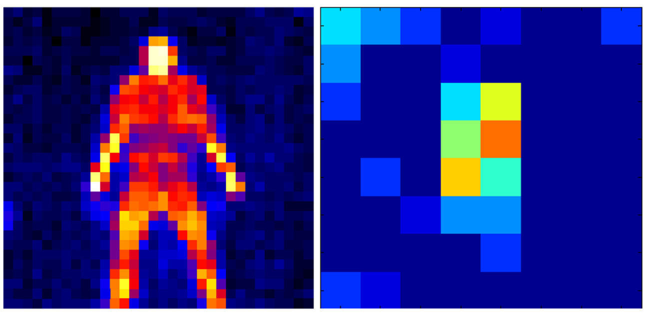
The purpose of this study is monitoring elderly people's daily life at home by recognizing their fundamental activities and detecting emergency situation like falling. In particular,

978-1-4673-6432-4/15/$31.00 ©2015 IEEE



(a) (b)

Fig. 1. Low resolution infrared array sensors. (a) the sensor that has 32×31 infrared detectors inside (HTPA32x31). (b) 25 cents coin and the sensor that has 8×8 infrared detectors inside (AMG8831).



(a) (b)

Fig. 2. Temperature distribution obtained from the low resolution infrared array sensor. Both captured the same situation. Bright color pixel indicates that there is a high temperature object (color scales are not same). (a) 32×31 pixels. (b) 8×8 pixels.

detection precision of emergency situation needs to be higher than that of any other activities.

This paper is organized as follows. Section II introduces infrared ray (IR) and infrared array sensors concisely. Section III presents the system model. Section IV details the proposed activity recognition method. Section V shows experimental results. Section VI concludes the paper.

II. Infrared Array Sensors

In general, all objects emit infrared rays (IR), and the stronger the infrared rays become, the higher the temperature of an object becomes. Infrared rays are similar to light but has a longer wavelength extending from the red edge of the visible spectrum at 0.7 μm to 1000 μm. Therefore, human cannot see it without special equipment. Typically, IR is divided into three parts: near-infrared rays, mid-infrared rays, and far-infrared rays, by the wavelength. Above all, far-infrared rays have characters like electromagnetic wave and can be detected using infrared sensors including low resolution infrared array sensors shown in Fig. 1.

Low resolution array sensors (hereinafter referred to as infrared array sensors) have $m \times n$ infrared detectors (or pixels) inside. From this, temperature distribution is obtained on a two dimensional area. The sensor is typically used for

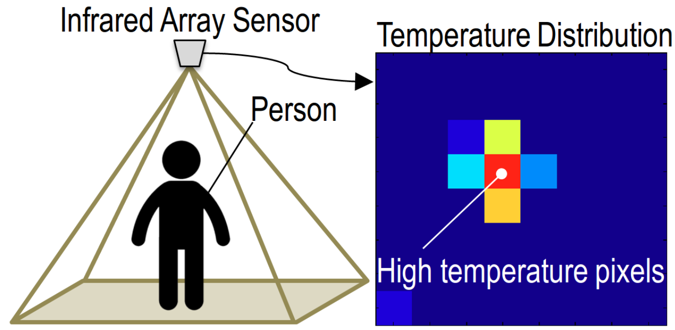


Fig. 3. System model of our proposed method. The infrared array sensor is installed on the ceiling and sends temperature distribution.

TABLE I
Definitions of Five Activities

Activity	Definition
No event	There is no person in the detection area.
Stopping	A person is standing, sitting, or lying still. Quiescent state.
Walking	A person is walking around in the detection area.
Sitting	Motion of sitting down. Posture after sitting down is classified as "Stopping".
Falling	Motion of falling down. Posture after falling down is classified as "Stopping".

high performance home appliance (microwave oven and air conditioner), energy savings in office [13] (air conditioning and lighting controls), digital signage and automatic door and elevator. In this paper, we define infrared thermography (also known as thermal imaging and thermal video) as the infrared sensors that can capture temperature distribution in high definition (e.g. 160×120, 320×240, and 640×512) and infrared array sensors as the infrared sensor that can capture temperature distribution in low resolution (e.g. 8×8, 16×4, and 32×31). Fig. 2 shows temperature distribution obtained from the infrared array sensor that has $m \times n$ pixels inside. As shown in this figure (particularly Fig. 2 (b)), we cannot distinguish individuals (privacy protection), because sensor's resolution is low. Furthermore, the sensor can detect a person in darkness by detecting IR. In recent years, the sensor has become small and cheap by technological advance (e.g. 8×8 infrared array sensor in Fig. 1 (b) costs approximately US$ 25 per piece).

III. System Model

In this paper, we propose an activity recognition method using a low resolution infrared array sensor as development of our previous work [12]. The sensor is installed on the ceiling as shown in Fig. 3, and activities taken in the detection area are classified into five classes: no event, stopping, walking, sitting and falling. Definitions of these activities are detailed in TABLE I. We assume that the sensor has $m \times n$ pixels

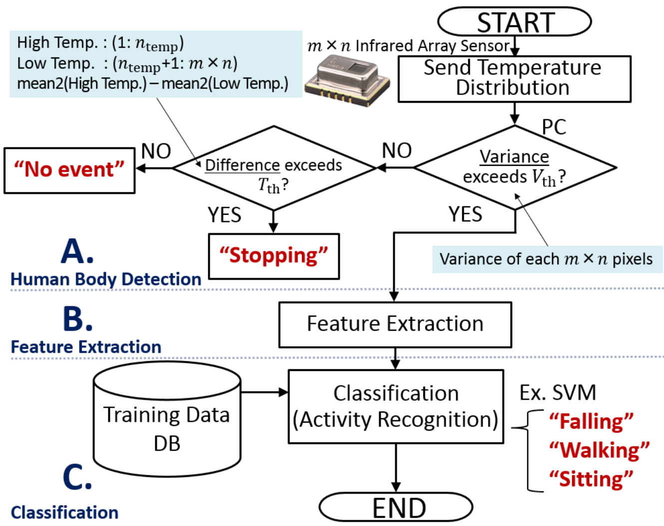


Fig. 4. The flowchart of proposed activity recognition algorithm. This algorithm can be divided into 3 steps: A. human body detection, B. feature extraction and C. classification.

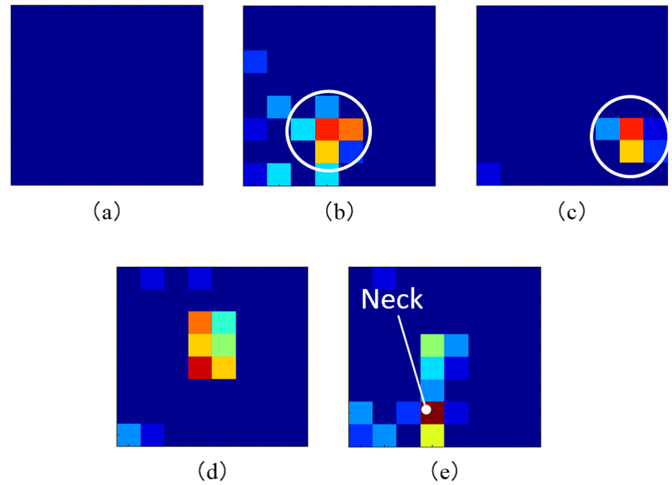


Fig. 5. Temperature distribution sent from 8×8 infrared array sensor. (a) No event, (b) Stopping, (c) Walking, (d) Sitting, and (e) Falling.

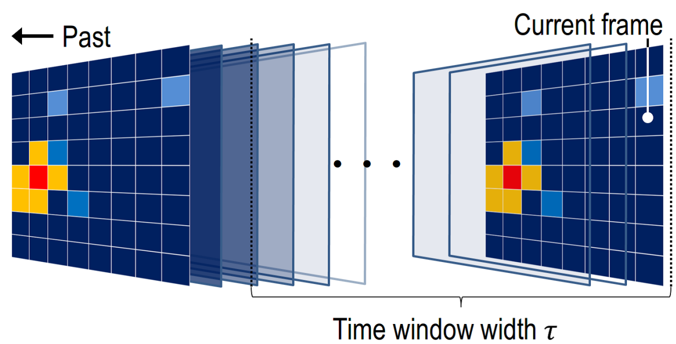


Fig. 6. Temperature variance of each pixel is calculated over τ frames. τ frames from the current frame are used. Time window width τ depends on sensor's frame rate.

(infrared detectors) inside, and the target is a single person who lives alone in ordinary home.

Consecutive temperature distribution images sent from the sensor is called frames. Frame rate depends on the sensor's specification. The frames are analyzed and classified by using algorithm detailed in next section.

IV. Proposed Activity Recognition Method

Fig. 4 indicates the flowchart of our proposed activity recognition algorithm. The algorithm is divided into three steps: human body detection, feature extraction, and classification. The general flow of the algorithm is as follows. First, in human body detection, activities are classified into two classes based on the temperature difference between a person and the background. Second, in feature extraction, four features focusing on human motions are extracted from temperature distribution. Finally, in classification, four features are compared with training data stored in database (DB) and classified as "Walking", "Sitting", or "Falling" by a classifier (e.g. support vector machine, nearest neighbor search or neural network).

A. Human Body Detection

In this step, activities are classified into two classes ("No event" and "Stopping"). At the beginning, temperature distribution of kth frame sent from $m \times n$ infrared array sensor (see Fig. 5) can be expressed as

$$\mathbf{T}_k = \begin{bmatrix} a_{11k} & a_{12k} & \cdots & a_{1nk} \\ a_{21k} & a_{22k} & \cdots & a_{2nk} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1k} & a_{m2k} & \cdots & a_{mnk} \end{bmatrix}, \quad (1)$$

where a_{ijk} is (i,j) temperature value of kth frame. We decide whether there is a moving person or not by calculating temperature variance of each pixel over τ frames. Temperature variance on the (i,j)th pixel of current frame is calculated as follows:

$$v_{ijk_c} = \frac{1}{\tau} \sum_{k=k_c-(\tau-1)}^{k_c} (a_{ijk} - \overline{a_{ijk_c}})^2, \quad k_c \geq \tau \quad (2)$$

where τ is time window width (detailed in Fig. 6), and k_c is current frame number. The (i,j) temperature average $\overline{a_{ijk_c}}$ of current frame is given by

$$\overline{a_{ijk_c}} = \frac{1}{\tau} \sum_{k=k_c-(\tau-1)}^{k_c} a_{ijk}, \quad k_c \geq \tau. \quad (3)$$

When one or more temperature variance v_{ijk_c} satisfy the following inequality (4), we consider there is a moving person in the detection area and move to step B, feature extraction.

$$v_{ijk_c} > V_{\mathrm{th}}. \quad (4)$$

The threshold V_{th} is decided so that variance v_{ijk_c} will not exceed the value when there is no person (v_{ijk_c} varies slightly

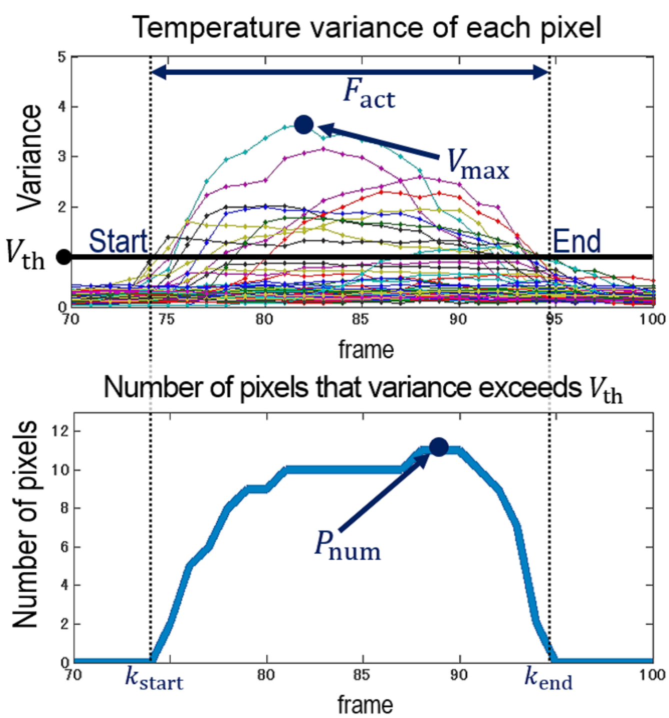


Fig. 7. Features: F_{act}, P_{num} and V_{max}.

even when there is no person).

Next, if we decide that there is no moving person (satisfy (5)), the state is classified as "No event" or "Stopping" by using temperature difference between a person and the background.

$$v_{ijk_c} \leq V_{\mathrm{th}}. \tag{5}$$

First, $\mathbf{T}_{ijk}$ is arranged in descending order, and the temperature average from maximum temperature to n_{temp}th highest temperature, and from $(n_{\mathrm{temp}}+1)$th highest temperature to minimum temperature ($m \times n$th highest temperature) are expressed as T_{p} and T_{b}, respectively. n_{temp} is decided based on the number of reacting pixels when a person is standing. Then, temperature difference between a person and the background is given by

$$T_{\mathrm{diff}} = T_{\mathrm{p}} - T_{\mathrm{b}}. \tag{6}$$

When T_{diff} satisfies

$$T_{\mathrm{diff}} \geq T_{\mathrm{th}} \tag{7}$$

over n_{cont} consecutive frames, the state is classified as "Stopping". In reverse, when T_{diff} satisfies

$$T_{\mathrm{diff}} < T_{\mathrm{th}} \tag{8}$$

over n_{cont} consecutive frames, the state is classified as "No event". The threshold T_{th} is decided based on temperature difference between a person and the background when the person is standing still.

Parameters mentioned above, τ, V_{th}, n_{temp}, T_{th} and n_{cont}, are decided in preliminary experiments, because they depend on sensor's performance (frame rate, temperature accuracy, and the number of pixels).

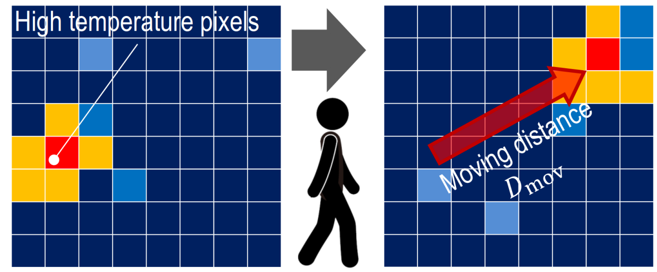


Fig. 8. The feature D_{mov} is the Euclidean distance. The left side indicates $(k_{\mathrm{start}} - k_{\mathrm{m}})$th frame and the right side indicates $(k_{\mathrm{end}} + k_{\mathrm{m}})$th frame.

B. Feature Extraction

If we decide that there is a moving person in the detection are, four features focusing on human motions are extracted from temperature distribution as our previous work [12].

1) Active Frame (F_{act}): When one ore more temperature variances v_{ijk} exceed the threshold V_{th}, the frame is regarded as the start of a motion k_{start}. Conversely, when all temperature variances v_{ijk} fall below the threshold V_{th}, the frame is regarded as the end of the motion k_{end} as shown in Fig. 7 (upper graph). This feature, F_{act} is given by

$$F_{\mathrm{act}} = k_{\mathrm{end}} - k_{\mathrm{start}} + 1. \tag{9}$$

This feature tends to be large when a motion that takes time occurs. For example, F_{act} of "Walking" tends to be larger than that of "Falling" and "Sitting".

2) Maximum Number of Reacting Pixels (P_{num}): The feature P_{num} is the maximum number of pixels that exceed the threshold V_{th} in the interval $[k_{\mathrm{start}}, k_{\mathrm{end}}]$ as shown in Fig. 7 (lower graph). P_{num} tends to be large when a motion like "Falling" occurs, because body superficial area observed by the sensor tends to be large.

3) Maximum Temperature Variance (V_{max}): The feature V_{max} is the maximum temperature variance in the interval $[k_{\mathrm{start}}, k_{\mathrm{end}}]$ as shown in Fig. 7 (upper graph). V_{max} tends to be large when a sudden motion occurs. For example, V_{max} of "Falling" tends to be larger than that of "Walking" and "Sitting".

4) Moving Distance (D_{mov}): In general, high temperature is observed around a head or a neck when temperature distribution is observed by the sensor installed on the ceiling. Accordingly, in this paper, we assume that there is a person in the position of the maximum temperature pixel. The feature D_{mov} is the Euclidean distance between person's position of $(k_{\mathrm{start}} - k_{\mathrm{m}})$th frame and that of $(k_{\mathrm{end}} + k_{\mathrm{m}})$th frame as shown in Fig. 8. The margin k_{m} depends on sensor's frame rate, and its value is decided from preliminary experiments. D_{mov} tends

TABLE II
SPECIFICATIONS OF THE SENSOR [14]

Item	Part Number / Performance
Part number	Grid-EYE AMG8831
Number of pixels	64 (Vertical 8 $\times$ Horizontal 8 Matrix)
Temperature output resolution	0.25 °C
Temperature range of measuring object	0 °C to 80 °C +32 °F to +176 °F
Temperature accuracy	Typical $\pm$2.5 °C $\pm$4.5 °F
Detection distance	Max. 5 m
Viewing angle	Typical 60°
Frame rate	Typical 10 frames/sec or 1 frame/sec

TABLE III
EXPERIMENTAL SPECIFICATIONS

Room	A	B
Ceiling height	2.57 m	2.75 m
Detection area	3.5 m$\times$3.5 m (12.25 m^2)	
Room temperature	23 °C	26 °C
Floor material	Carpet	Synthetic resin
Subjects	A, B, C, D, E, F	A, C
Activities	No event, Stopping, Walking, Sitting, Falling	
Classifier	Support Vector Machine	
Library	LibSVM [16]	
Kernel	Radial Basis Function	
Test data	Subjects: A and B.	
Training data	Subjects: C, D, E and F.	
V_{th}	1.0 °C^2	
T_{th}	1.860171 °C	
τ	20	
n_{temp}	5	
n_{cont}	3	
k_{m}	10	

to be large when a motion witch takes long distance occurs. For example, D_{mov} of "Falling" tends to be larger than that of "Sitting".

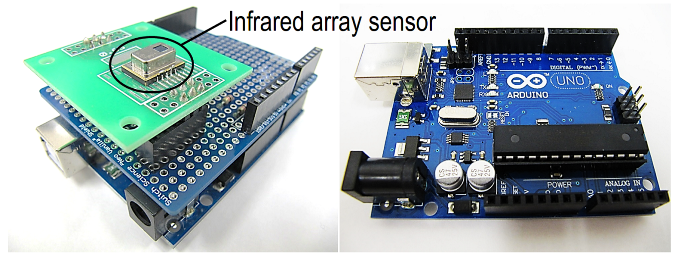
C. Classification

Finally, in this step, extracted features (test data) are compared with training data stored in database (DB). Therefore, we need to conduct preliminary experiments and store enough training data in advance. As the classifier, for example, support vector machine, nearest neighbor search or neural network are used (support vector machine was used in the experiments).

V. EXPERIMENTS

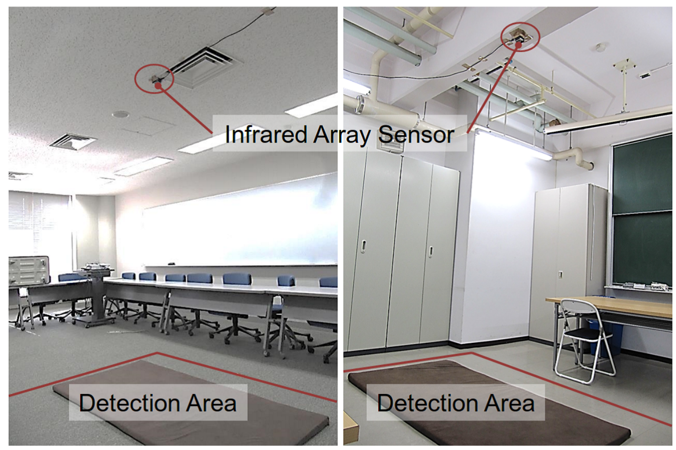
A. Device

We use the infrared array sensor called "Grid-EYE" [14] in the experiments. Specifications of this sensor are shown in TABLE II. The sensor is attached to the microcomputer board, Arduino UNO Rev3 [15], and temperature distribution is sent to a PC via USB port as shown in Fig. 9. The received



(a) (b)

Fig. 9. The sensing device used in experiments. (a) the sensing device is composed of an infrared array sensor (Grid-EYE) and a microcomputer board (Arduino UNO Rev3). (b) Arduino UNO Rev3.



(a) (b)

Fig. 10. Experimental environments. (a) room A and (b) room B. Ceiling height, temperature, and floor material of room A are different from those of room B.

temperature distribution is analyzed and classified into five activities by support vector machine where we use LibSVM [16].

B. Experimental Specifications

The experiments were carried out in two rooms shown in Fig. 10. Experimental specifications are shown in TABLE III, where test data include 30 fallings, 20 sittings and 27 walkings, and training data include 50 fallings, 30 sittings and 45 walkings. Recognition accuracy is calculated as follows:

$$\frac{\text{The number of correctly classified frames}}{\text{The number of total frames}} \times 100\ [\%]. \quad (10)$$

C. Classification Results

Classification results are shown in TABLE IV. Except for "Siting", our proposed method achieved high recognition accuracy of more than 94 %. If we express the results as the number of activity times, 30/30 fallings, 17/20 sittings and 26/27 walkings were classified correctly. When subjects sat down forcefully, the activities tended to be misclassified as

TABLE IV
CONFUSION MATRIX OF CLASSIFICATION RESULTS

		Classified Class [frame]				
		No event	Stopping	Falling	Sitting	Walking
Actual Class	No event	2893	0	0	0	0
	Stopping	847	15590	0	0	0
	Falling	0	0	628	0	0
	Sitting	0	0	95	348	0
	Walking	0	2	0	0	3042
Accuracy		100.0 %	94.8 %	100.0 %	78.6 %	99.9 %

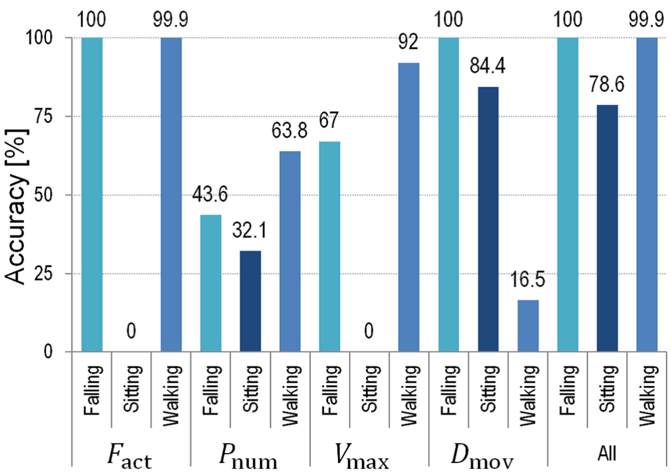


Fig. 11. Results of feature evaluation. Classification results of "Falling", "Sitting", and "Walking" when we use each feature.

"Falling", because the features (particularly F_{act}) took values similar to that for "Falling".

D. Feature Evaluation

Results of feature evaluation are shown in Fig. 11. In this figure, "All" is the combination of all features: F_{act}, P_{num}, V_{max}, and D_{mov}. "Sitting" recognition results by F_{act} and V_{max} became 0 %. F_{act} and D_{mov} have an effect on "Falling" recognition, because time of the falling motion and the moving distance (the moving distance of a head or a neck) are more characteristic than the other activities. F_{act} of "Walking" tends to take quite larger value than the other activities. To enhance "Sitting" recognition accuracy, we need to find new features or reconsider the placement of the sensor (e.g. we can get height information by installing the sensor on the wall).

VI. CONCLUSION

In this paper, we proposed activity recognition using low resolution infrared array sensor. From the viewpoint of general versatility (available in darkness), cost, size, privacy (low resolution) and availability (commercial off-the-shelf), the sensor can be suitable for elderly people monitoring. We have proposed a fall detection method using the sensor to inform an unexpected falling in previous work. In addition to falling detection, knowing daily fundamental activities of elderly people is also important to prevent future accidents. Therefore, in our new proposed method, we enabled five activities recognition, "No event", "Stopping", "Walking", "Sitting", and "Falling" (emergency situation), by using temperature difference between a person and the background, and four features focusing on human motions. As a result of experiments, our proposed method achieved "No event", "Stopping", "Walking", and "Sitting" recognition accuracy of 100 %, 94.8 %, 99.9 %, and 78.6 %, respectively. In particular, 100 % accuracy of "Falling" (emergency situation) recognition was achieved.

To enhance "Sitting" recognition accuracy, we might need new features and placement of the sensor for improvements. Classifying the states when a person is stopping (No event) can also be the key to improve recognition accuracy. We hope that our proposed method helps elderly people's daily life.

REFERENCES

[1] United Nations, Department of Economic and Social Affairs, Population Division. (2013). *World Population Aging 2013* [Online]. Available: http://www.un.org/esa/socdev/documents/ageing/Data/WorldPopulationAgeingReport2013.pdf
[2] Cabinet Office, Government of Japan. (2013). *Annual report on the aging society: 2013 (Summary)* [Online]. Available: http://www8.cao.go.jp/kourei/english/annualreport/2013/2013pdf_e.html
[3] M. Fahim, I. Fatima, S. Lee, and Y. T. Park, "EFM: Evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer," *Appl. Intell.*, vol. 39, no. 3, pp. 475–488, Oct. 2013.
[4] X. Su, H. H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua Sci. Technol.*, vol. 19, no. 3, pp. 235–249, June 2014.
[5] M. Zhang and A. A. Sawchuk, "Human daily activity recognition with sparse representation using wearable sensors," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 3, pp. 553–560, May 2013.
[6] A. M. Khan, Y. K. Lee, S. Y. Lee, and T. S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1166–1172, Sept. 2010.
[7] J. Cheng, O. Amft, G. Bahle, and P. Lukowicz, "Designing sensitive wearable capacitive sensors for activity recognition," *IEEE Sensors J.*, vol. 13, no. 10, pp. 3935–3947, Oct. 2013.
[8] J. Hong and T. Ohtsuki, "Array sensor: Monitoring using wireless signals," in *Proc. Int. Conf. ICT Convergence (ICTC)*, pp. 757–758, Oct. 2013.
[9] W. Lin, M. T. Sun, R. Poovandran, and Z. Zhang, "Human activity recognition for video surveillance," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, pp. 2737–2740, May 2008.
[10] J. Hong, S. Tomii, and T. Ohtsuki, "Cooperative fall detection using Doppler radar and array sensor," in *Proc. IEEE 24th Int. Symp. Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 3492–3496, Sept. 2013.
[11] S. Tao, M. Kudo, H. Nonaka, and J. Toyama, "Camera view usage of binary infrared sensors for activity recognition," in *Proc. Int. Conf. Pattern Recognition (ICPR)*, pp. 1759–1762, Nov. 2012.
[12] S. Mashiyama, J. Hong, and T. Ohtsuki, "A fall detection system using low resolution infrared array sensor," in *Proc. IEEE 25th Int. Symp. Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 2114–2118, Sept. 2014.
[13] L. I. L. Gonzalez, M. Troost, and O. Amft, "Using a thermopile matrix sensor to recognize energy-related activities in offices," in *Proc. The 3rd Int. Conf. Sustainable Energy Information Technology (SEIT 2013)*, vol. 19, pp. 678–685, June 2013.
[14] Panasonic Corporation. *Infrared array sensor Grid-EYE* [Online]. Available: http://www3.panasonic.biz/ac/e/control/sensor/infrared/grid-eye/index.jsp
[15] Arduino. *Arduino UNO Rev3 (arduino store)* [Online]. Available: http://store.arduino.cc/product/A000066
[16] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Int. Syst. and Tech. (TIST)*, vol. 2, no. 3, pp. 1–27, Apr. 2011.