Indonesian Journal of Electrical Engineering and Computer Science
Vol. 5, No. 3, March 2017, pp. 549 ~ 555
DOI: 10.11591/ijeecs.v5.i3.pp549-555

Human Presence Recognition in a Closed Space by Using Cost-effective CO2 Sensor and the Information Gain Processing Method

Kimio Oguchi*[1], Ryoya Ozawa[2]
[1]Information Networking Laboratory, Graduate School of Science and Technology, Seikei University
[1,2]Information Networking Laboratory, Faculty of Science and Technology, Seikei University,
3-3-1 KichijojiKitamachi, Musashino, 180-8633 Tokyo, Japan, Tel: +81-422-37-3728
*Corresponding author, e-mail: oguchi@st.seikei.ac.jp

Abstract

The recent rapid progress in ICT technologies such as smart/intelligent sensor devices, broadband / ubiquitous networks, and Internet of everything (IoT) has advanced the penetration of sensor networks and their applications. The requirements of human daily life, security, energy efficiency, safety, comfort, and ecological, can be achieved with the help of these networks and applications. Traditionally, if we want some information on, for example, environment status, a variety of dedicated sensors is needed. This will increase the number of sensors installed and thus system cost, sensor data traffic loads, and installation difficulty. Therefore, we need to find redundancies in the captured information or interpret the semantics captured by non-dedicated sensors to reduce sensor network overheads. This paper clarifies the feasibility of recognizing human presence in a space by processing information captured by other than dedicated sensors. It proposes a method and implements it as a cost-effective prototype sensor network for a university library. This method processes CO2 concentration, originally designed to check environment status. In the experiment, training data is captured with none, one, or two subjects. The information gain (IG) method is applied to the resulting data, to set thresholds and thus judge the number of people. Human presence (none, one or two people) is accurately recognized from the CO2 concentration data. The experiments clarify that a CO2 sensor in set in a small room to check environment status can recognize the number of humans in the room with more than 70 % accuracy. This eliminates the need for an extra sensor, which reduces sensor network cost.

Keywords: *CO2 sensor, Human presence, Information gain, Information redundancy*

1. Introduction

Recent rapid advances in ICT (Information and Communications Technology) has yielded a new wide range of systems such as CPS (Cyber Physical System) and a variety of entities such as sensor devices/modules, sensor networks, signal/data transfer networks, and data analyzers (Big data or machine learning). These CPSs are now indispensable for our daily life as they offer generally security, efficiency, safety, comfort, and energy friendliness (green). Therefore, active research in CPS-related area e.g. devices, systems, and applications is being conducted everywhere.

Our laboratory has been researching several CPS related systems and applications that mainly focus on the vital human information captured by wearable sensor(s) [1, 2] and environmental information capture [3, 4]. In our research on environmental information capture, we found that human presence recognition is one of the key pieces of information needed in managing the environment, and ensuring security.

In general, if we want some information on, for example environment status, dedicated sensors are needed. Cameras are clearly a solution for human presence recognition, but they come with the costs of privacy loss, data storage capacity, and high transfer bandwidth for data.

Using dedicated sensors will increase the number of sensors and thus system cost, sensor data traffic loads, and installation difficulty. Therefore, finding some redundancy in the captured information, or interpreting the semantics captured by non-dedicated sensors will reduce sensor network overheads.

Received November 10, 2016; Revised January 30, 2017; Accepted February 16, 2017

This paper clarifies the feasibility of recognizing human presence in small spaces by processing information captured without dedicated sensors. It introduces and verifies the proposed CO2 method, originally intended to check environment status, as implemented in a cost-effective prototype CO2 sensor network for a university library.

This paper is structured as follows: in Section 2, library requirements in terms of environmental management are first overviewed. Then, the experimental method is described. The information gain (IG) method for judgement is also described. In Section 3, the results are discussed, and a conclusion and future works are given in Section 4.

Parts of the research background, the general requirements for the university library described in Section 2 have already been presented by the authors' laboratory [5, 6]. However, the sensor chip used and the human presence recognition procedure, the IG method, differ from those of the previous papers.

2. Research Method

2.1. Requirements for University Library

Generally speaking, environmental requirements for libraries differ greatly from those for the office/home as they are heterogeneous spaces where human(s), books and other stored materials co-exist.

Our survey on the environmental systems in 28 large libraries in Japan that covered the current situation with regard to building structures and air conditioning systems identified the following important issues [7].

1. Most libraries building have a large atrium with at least one glass wall and a high roof,
2. In most libraries, the managers have to manually fine-tune the air conditioning system,
3. Most libraries do not have a sofisticated IT infrastructure e.g. sensor network (SN),
4. The air conditioning system must be able to create a comfortable environment for both users and books, and
5. Most users and managers are dissatisfied with their current air-conditioning systems.

In addition to those requirements, the Seikei University library needs to check whether small personal-use rooms are actually occupied, regardless of its reservation state. The reason is sometimes a booked room is not used. Lower system cost and strong privacy concerns are other issues that need to be addressed.

To resolve these issues and to create smart/intelligent library environments for both users and books, the use of Information Communication Technologies (ICTs), especially SNs is essential with minimum sensor installation.

2.2. General System Configuration of the Proposed Method

Basic configuration of the proposed CO2 method is depicted in Figure 1. Even though just a single CO2 sensor is discussed in this paper, a transmitter (Tx) generally has one or more sensors that capture several kinds of environmental information. The processing unit (PU) gathers the data from a set of sensors, and transfers them to the next node. Sometimes the PU performs data reduction in terms of capacity, or analyzes them for triggering alerts/actions.

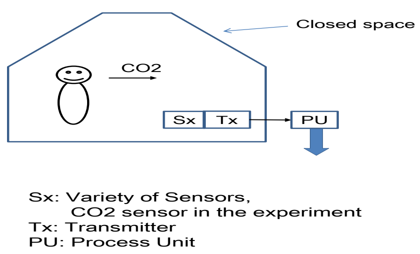


Figure 1. Basic configuration of the proposed CO2 method.

Experimental procedures are as follows:
1. Collect CO2 data (training data) in the laboratory with none, 1, 2 and 3 people,
2. Determine threshold value for the IG method to judge whether a room is occupied,
3. Collect CO2 data (test data) in a small room in the university library,
4. Estimate occupation state by applying the threshold, and
5. Discuss accuracy.

2.3. Judgement Method

In order to judge whether or not a small room is occupied, the notion of relative information gain (RIG), or relative value of information gain (IG) is used in the experiment. IG is a machine learning method and a measure of the amount of uncertainty of the input of a system given the value of the output, simply the expected reduction in entropy; details are found in [8].

The *IG* between two random variables *x* and *y* is defined as:

$$IG(y, x) = H(y) - H(y|x), \quad (1)$$

Where $H(y)$ is the general entropy equivalent to the inherent uncertainty of random value *y,* and is given as:

$$H(y) = - \sum_{i=1}^{n} p(y_i) \log_2 p(y_i). \quad (2)$$

Here, *n* equals the total number of values. $H(y|x)$ is the entropy of *y* when variable *x* is known.

The relative information gain, *RIG* is defined as:

$$RIG(y, x) = IG(y, x) / H(y). \quad (3)$$

After obtaining training data, the CO2 ratio over time is evaluated. Several CO2 reference values (RVs) are compared by using IG and RIG. Here, base RV is set by using the general CO2 increase ratio (IR) as detailed in the next subsection. Most probable RV, the one that has highest RIG value, is selected.

2.4. General CO2 Increase Ratio

CO2 is being created in large quantities and its concentration is continuously increasing that is one reason for global warming [9]. About 400 ppm is the current value (this is called the baseline hereafter).

The increase in CO2 concentration by human respiration in a closed space (INC_{CO2}) is given simply as:

$$INC_{CO2} = Vol_{exhaled} \times Con_{CO2} \times n_{breath} / Vol_{space}, \quad (4)$$

Where $Vol_{exhaled}$ is the volume of each exhaled breath, Con_{CO2} the CO2 concentration in an exhaled breath, n_{breath} the respiratory rate per minute, Vol_{space} the volume of the closed space. Typical values for adults are $Vol_{exhaled}$ = 0.5 L, Con_{CO2} = 4 %, n_{breath} = 15 times per min. Thus (4) is rewritten as:

$$INC_{CO2} = 0.3 / Vol_{space} \ [\mathrm{min}^{-1}]. \quad (4a)$$

As the volume of the closed space used in the experiment was 4,540 L (details below in 2.6), equation (4a) gives the INC_{CO2} value of 66.7 ppm/min. This means that if a person is present in the space, CO2 concentration will typically increase at the rate of 66.7 ppm per minute from the baseline. This value is used for base RV (bRV) as described in 2.3.

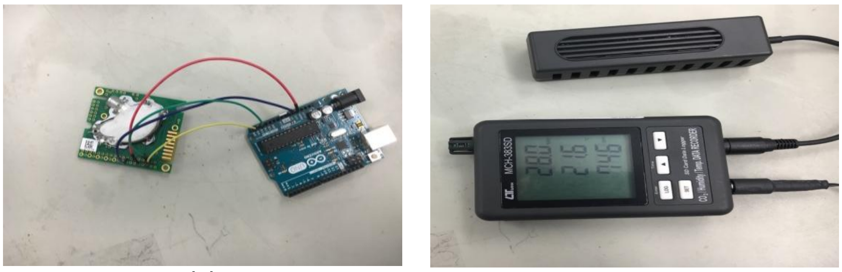
2.5. Sensor Configuration

Two sensor modules used in the experiments are i) a prototype unit consisting of a K30 sensor chip (CO2 Meter [10]) and an Arduino Uno [11] as a signal processor (e.g. analog to digital conversion) and transmitter (Tx) with USB connection, and, ii) commercially available unit that is a digital CO2 meter (MCH-383SD, Mother Tool [12]). The PU is a Laptop PC (Lenovo

G580; Win8; Celeron 1000M 1.8 GHz; 4 GB). Basic characteristics of these sensors are given in Table 1. Photographs are given in Figure 2 (a) and (b).

Table 1. Basic Characteristics of CO2 Sensors Used

Item	Prototype: K30 chip (Module-K30)	MCH-383SD (Module-383)
Measureable CO2 range [ppm]	0 – 5,000	0 – 4,000
Operating temperature [degree C]	0 - 50	0 – 50
Size (L x W x D) [mm]	51 x 57 x 12.5	132 x 38 x 32**
Operating voltage [V]	4.5 – 14*	9
Note	* 5.0 V with Arduino	**Sensor module only



(a) (b)

Figure 2. (a) Prototype; K30 chip (left) and Arduino module (Module-K30), and (b) commercially available MCH-383SD (Module-383) where the upper part is sensor module

(a) (b)

Figure 3. (a) Closed space for training data capture in the laboratory, and (b) a small room in the library. Photos do not have the same scale

2.6. Spaces in the Experiment and Subjects

The closed space used for training data in the laboratory is a walk-in greenhouse with durable clear plastic cover and a roll up door; its dimensions are roughly 1.2 x 1.9 x 1.9 m (W x L x H); a photo is given in Figure 3(a). This was assembled in the laboratory for this test.

The small room in the library has dimensions of about 1.6 x 1.8 x 3.0 m (W x L x H); a photo is given in Figure 3(b). All subjects were young male volunteers with ages of 22 or 23y.

3. Results and Analysis

3.1. Training Data

Training data were collected in the closed space in the laboratory with none, 1, 2, and 3 people present as indicated in Figure 3(a). Starting with none, people entered the closed space one after the other at 5-minute intervals. Thirteen trials in total were conducted.

According to the data collected, CO_2 concentration increased with the addition of each person. However, the increase was delayed by about 2 min due to the delay in sensor response.

The increase ratio of each trial was calculated by linear approximation. Table 2 summarizes the increase ratio of CO_2 concentration for none, 1, 2 and 3 people.

3.2. Reference Value (RV)

Most probable RV is selected by comparing IG and RIG. Here, the base RV (bRV) is set to 66.7 ppm/min as discussed in 2.4. Seventeen RV ranges were examined starting with bRV +/- 5 ppm, i.e. 61.7 ≤ RV ≤ 71.7 and extending in +/- 5 ppm increments.

IG and RIG are calculated as follows for the first RV range, 61.7 ≤ RV ≤ 71.7. Here, assume S is a collection containing 52 examples as indicated in Table 2, and O is the number of occupants. The notation [2+, 50-] means 2 positive (matched) and 50 negative (unmatched) examples were present in the 52 examples included in S [8].

$$S_{(61.7 \le RV \le 71.7)} = [2+, 50-] \tag{5}$$
$$H(S) = -(2/52)\log_2(2/52) - (50/52)\log_2(50/52)$$
$$= 0.2351$$

$$S_{(O=1,\ 61.7 \le RV \le 71.7)} = [2+, 11-] \tag{6}$$
$$H(S) = -(2/13)\log_2(2/13) - (11/13)\log_2(11/13)$$
$$= 0.6194$$

$$S_{(O \ne 1,\ 61.7 \le RV \le 71.7)} = [0, 39-] \tag{7}$$
$$H(S) = -(0/39)\log_2(0/39) - (39/39)\log_2(39/39)$$
$$= 0$$

Therefore, IG and RIG for the first RV range are given as:

$$IG(S, RV) = H(S) - \Sigma_{v \in \{O=1,\ O \ne 1\}} |S_v|\, H(S_v) / |S| \tag{8}$$
$$= 0.2351 - 0.1548 - 0$$
$$= 0.0803$$

$$RIG(S, RV) = IG(S, RV) / H(S) \tag{9}$$
$$= 0.0803 / 0.235$$
$$= 34.2\ (\%)$$

Table 2. Increase Ratio of CO_2 Concentration for None, 1, 2 and 3 Person Occupancy

Data No	Number of occupants			
	None	1	2	3
1	1.938	54.92	139.1	272.0
2	-3.126	72.96	128.9	161.3
3	-6.58	97.06	155.6	257.9
4	1.443	119.3	151	228.2
5	11.05	76.22	188	270.1
6	35.05	66.41	126.4	219.5
7	5.323	47.12	194.9	214.5
8	-14.52	106	181.9	196
9	10.95	63.49	191.5	232.2
10	2.492	111.1	169.9	129.1
11	-2.368	103.5	136.1	146.3
12	-7.385	91.46	100.5	226.1
13	-13.91	87.59	204.7	242.8
Average	1.57	84.39	159.1	215.1

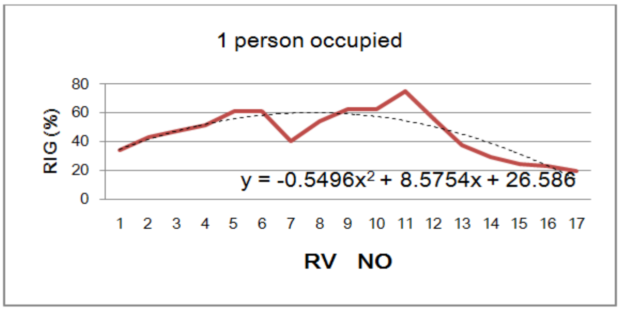


Figure 4. RIG values for all RV ranges (solid brown line), together with its approximation curve (dashed line)

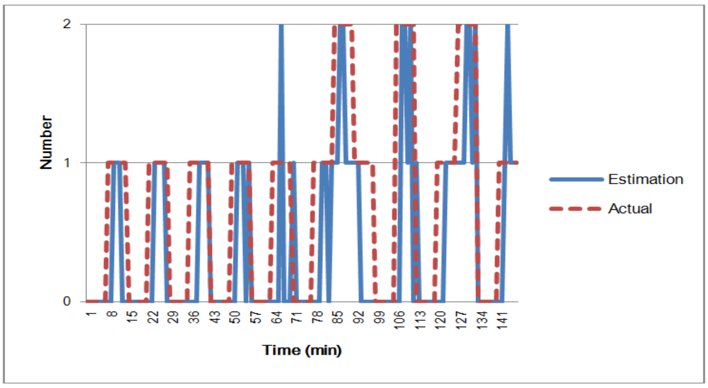


Figure 5. Test data collected in the university library. Dashed brown line corresponds to actual occupancy and solid blue line to estimated result

This means that the assumption of “measured CO2 concentration range 61.7 to 71.7 ppm implies a single occupant” is 34.2 % correct.

The second and following RV ranges are also calculated in the same manner. Figure 4 plots RIG values for all 17 RV ranges evaluated. Among these RV ranges, 36.7 ≤ RV ≤ 96.7 was selected as the threshold for estimation of single occupant, and used to verify the test data.

3.3. Test Data Collected in the Small Room in the University Library

Test data was collected in the small room in the university library with none, 1 and 2 people as indicated in Figure 3(b). People randomly entered/left the room at 7 min intervals. Total time period was 147 min. Details are depicted by the dashed line in Figure 5 where the vertical axis is the number of occupants and the horizontal axis is time.

3.4. Estimation Results from Test Data

The RV range used for analyzing training data was 36.7 ≤ RV ≤ 96.7. However, as the small room in the university library is larger, a new range is needed. By applying the volume ratio of 4,540 /8,640 = 0.525, the adjusted RV range used for assessing the test data was taken to be 19.3 ≤ RV ≤ 50.1.

Solid blue line in Figure 5 depicts estimated values. The estimation judged the occupancy rate to be zero when the value is smaller than RV. It was taken to be two when the value is larger than RV. Judgment was done at every minute. Therefore, total number of judgments was 147 for the time period. Most judgments were accurate. However, the system exhibits some delay in capturing increases in the occupancy rate, e.g. 0 to 1, and 1 to 2 people. This might be due to the sensor response as also observed in subsection 3.1. 97 of the 147 estimation results were accurate, so that accuracy was 66.0 %.

3.5. Enhancing Estimation Accuracy

As the slow sensor response might yield lower estimation accuracy, several window widths that were not used for estimation were considered. With enlarging the window's width not used by 30 sec each, estimation accuracy increased, until reaching its maximum of 70.0 % with 150 sec window. Thereafter, it decreased. This implies the sensor module used (Module-K30) for the small room in the library has a response time for CO2 sensing of 150 sec and eliminating the window would increase estimation accuracy.

4. Conclusion and Future Work

Libraries are heterogeneous spaces for humans and books. In order to create smart/intelligent environments for libraries with ICTs, the conventional approach with sensor networks using dedicated sensors, increases the number of sensors, system cost, sensor data traffic loads, and installation difficulty. Therefore, finding some redundancy in the captured information, or interpreting the semantics captured by non-dedicated sensors can reduce the number of sensors, sensor network loads and system cost.

This paper proposed a method to recognize human presence in spaces by analyzing CO2 data with the information gain method; its feasibility was experimentally verified. Experiments showed that human presence can be recognized with an accuracy of about 70 % for occupancy of none, 1 and 2 people, without any dedicated sensors. This accuracy may or may not satisfy the requirement. It depends on what system/application will be used, together with considering e.g. total system cost.

Future works include integrating other methods e.g. humidity change, for higher accuracy without any dedicated sensors, and verifying the performance of human recognition and original environmental information capture as well. Continuous information capture in the field is also planned.

References

[1] Oguchi K. Content Transfer and Supporting Technologies in a Home Environment over Next Generation Convergence Home Network - From Vital Information Transfer to Broadband Content Transfer. *International Journal of Digital Content Technology and its application (IDCTA)*. 2009; 3(3): 124-135.

[2] Oguchi K, Enomoto Y, Terada S, Hanawa D. *Sensor-based e-Healthcare in a next generation convergence home network*. 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies. Barcelona. 2011.

[3] Hanawa D, Yamamoto H, Koide T, Shoji T, Suzuki K, Yamakawa S, Watanabe I, Oguchi K. Sensor networks for creating comfortable environments in university libraries. *Journal of CIE*. 2010; 33(5): 645-652.

[4] Igarashi K, Hanawa D, Oguchi K. *Estimation for spatial distribution of temperature and humidity by using sensor networks.* 34th International Conf. on Telecommunications and Signal Processing. Budapest. 2011: 227-230.

[5] Oguchi K, Maruta S, Hanawa D. *Human positioning estimation method using received signal strength indicator (RSSI) in a wireless sensor network*. Procedia Computer Science. 2014; 34: 126-132.

[6] Oguchi K, Tomono T, Hanawa D. *Multiple use of the information captured by cost-effective sensor networks in university library.* International Conference on Ubiquitous and Future Networks. Sapporo. 2015: 749-753.

[7] Yamamoto H, Suzuki K, Koide T, Yamakawa S, Hanawa D, Watanabe I, Oguchi K. *Investigation on and air-conditioning system of the library in Japan.* 10th International Conference on Computers, Communications and System. Daegu. 2009: 112-115.

[8] Mitchell TM. Machine Learning. New York: McGraw-Hill. 1997.

[9] IPCC (International Panel on Climate Change). *Fourth assessment report: Climate change*. 2007.

[10] SenseAir. CO2 Engine®K30 STA. SAKAKI Corp. Tokyo: 2016.

[11] Banzi M. Getting started with Arduino. 2nd Edition. O'Reilly. 2011.

[12] Lutron. CO2/Humidity/Temp. Data Recorder Model: MCH-383SD. http:// http://www.sunwe.com.tw/lutron/MCH-383SDeop.pdf