

Evaluating People's Attention in the Real World

Niwat Thepvilojanapong^{1,7} Yasunori Yakiyama² Oru Mihirogi³ Masayuki Iwai⁴
Kazunori Umeda⁵ Yoshito Tobe^{1,7} Ryosuke Shibasaki⁶

¹School of Science and Technology for Future Life, Tokyo Denki University, Tokyo, Japan

Tel: +81-3-5281-2309 E-mail: {wat, yoshito_tobe}@osoite.jp

²School of Advanced Science and Technology, Tokyo Denki University, Tokyo, Japan (E-mail: yakly@u-netlab.jp)

³Department of Information Systems and Multimedia Design, Tokyo Denki University, Tokyo, Japan

⁴Institute of Industrial Science, University of Tokyo, Tokyo, Japan (Email: masa@iis.u-tokyo.ac.jp)

⁵School of Science and Engineering, Chuo University, Tokyo, Japan (Email: umeda@mech.chuo-u.ac.jp)

⁶Center for Spatial Information Science, University of Tokyo, Tokyo, Japan (Email: shiba@csis.u-tokyo.ac.jp)

⁷Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Japan

Abstract: Although human activities in the World Wide Web are rapidly increasing due to the advent of many online services and applications, we still need to appraise how things such as a merchandise in a store or pictures in a museum receive attention in the real world. To measure people's attention in the physical world, we propose a *Sensor of Physical-world Attention using Laser scanning (SPAL)*. It is challenging to use a laser scanner because it provides only front-side circumference of any detected objects in a measurement area. Unlike cameras, a laser scanner poses no privacy problem because it does not recognize and record an individual. SPAL includes many important factors when calculating people's attention, i.e., lingering time, direction of people, distance between people and a target object. To obtain such information for calculation, we develop three processing modules to extract information from raw data measured by a laser scanner. We define two attention metrics and two measurement models to compute people's attention. To validate the proposed system, we implemented a prototype of SPAL and conducted experiments in the real-world environment. The results show that the proposed system is a good candidate for determining people's attention.

Keywords: People's attention, SPAL, laser scanner, laser range finder, human detection, direction detection, attention metrics, attention value, attention degree, experiments, evaluation.

1. INTRODUCTION

Laser scanners and image sensors have been used for measuring human behavior [11, 12, 15]. Most of previous works focus on the detection or tracking of humans. The measurements of such works are applied to traffic management and security systems [5, 13]. However, a measurement of people's attention has not been studied sufficiently. People's attention indicates how an object attracts the interest of people. We believe that a measurement of people's attention benefits our society as a metric for many purposes such as a marketing strategy and attaining comfortable urban activity. In the case of marketing strategy, we can place objects at the right places in an exhibition hall or a store in order to attract higher interest of people.

This paper proposes a *Sensor of Physical-world Attention using Laser scanning (SPAL)* system to measure people's attention from their behavior in the real world. Our system focuses on the attention received by target objects in various environments such as conventional halls, museums, department stores, shops, etc. Unlike stereo cameras, a laser scanner does not record images or recognize an individual, thus a privacy issue is not a problem for our system.

There are many challenges to use a laser scanner because it provides only front-side circumference of any detected objects in a measurement area. To measure people's attention, we develop a new technique to detect an

individual human by using only one laser scanner. Our technique is able to extract a human from a group of humans. Since a human with high interest will stop to watch a target object, we propose an algorithm to determine stationary people from all detected humans. In addition, to identify whether an extracted human who stays close to a target object really pays attention to the object, we also propose an algorithm to detect the direction of a human in comparison with the position of target object. After attaining the information of people who pay attention to a target object, the system calculates real-world attention levels by using two proposed models: *Object-based Attention (OA)* and *Distance-weighted Attention (DA)* models. We also define two metrics, *value* and *degree* of real-world attention levels, to indicate and evaluate people's attention. We conducted experiments in the real world to validate the feasibility of SPAL system including human detection, measurement models, and attention metrics. The experimental results show high accuracy of attention levels measured by the proposed system.

We briefly summarize our contributions as follows:

- We develop an algorithm to detect an individual human who pay attention to an object. To the best of our knowledge, this is the first attempt to detect the direction of a human by using only one laser scanner.
- We propose two measurement models to compute people's attention.
- We define two kinds of metrics to evaluate people's attention.

- We conducted experiments to evaluate the proposed system.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 proposes algorithms to detect an individual human and his/her direction towards a target object. Section 4 defines two attention metrics and proposes two measurement models for calculating real-world attention. The implementation of SPAL system and experimental results are reported in Section 5. Finally, we conclude this paper and describe future works in Section 6.

2. RELATED WORK

In computer science, *attention* is often referred to as *visual attention* in image processing when recognizing objects. Ma et al. [6] defines a user attention model which estimates attentions that viewers may pay to video contents. They model both visual and aural attentions corresponding to features extracted from video frames. For the same purpose, many visual attention models have been proposed for video summarization [4, 7, 8]. In addition, a human-robot interaction system also requires attention mechanisms to comprehend a situation [2]. Our focus is different from the above visual attention studies. We define *attention* as interest of people in the real world.

Our system exploits techniques of people tracking, trajectory detection, pedestrian counting, and crowd flow detection as its fundamental basis. Zhao et al. proposed to scan feet of pedestrians by a laser scanner or multiple laser scanners, and then analyzed walking trajectory based on a pedestrian model [3, 12, 14]. The system has a high robustness and yields a high accuracy even when people crowd the measured area. They applied their system to visualize the flows of passengers in railway stations [9]. In contrast, we use a laser scanner to scan the chest level of pedestrians, and be able to detect the direction of an individual human towards a target object.

In an intelligent transportation system, a system to recognize pedestrians provides useful information for drivers in order to achieve safe driving [1]. A stereo camera is often used to recognize pedestrians from images [11, 15].

As discussed above, laser scanners and stereo cameras are two major devices for tracking and detecting the flow of people. Although a stereo camera is able to measure a distance from an object, its accuracy is lower than that of laser scanners. The usage of stereo camera is limited by environmental phenomena such as the brightness of measurement area, and it also poses a privacy problem when using in public spaces. Therefore, we consider using a laser scanner which has many challenging issues in our work.

3. ATTENTION DETECTION SYSTEM

This section presents the algorithms used by the SPAL system to detect people who pay attention to a target object. The algorithm consists of three main steps, i.e., detecting human objects, extracting stationary humans, and

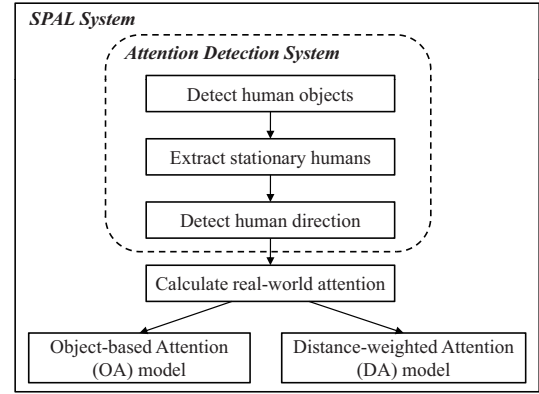


Fig. 1 The processing flow of SPAL system.

calculating human direction (dashed round-corner rectangle in Fig. 1).

3.1 Detection of Human Object

The system performs a step of detecting human objects by using HUMANDETECTION algorithm (Algorithm 1). The algorithm is able to extract a human from a group of human objects. First, the system generates background data ($BGData$) by scanning a measurement area while no one is there (Line 11). The background data is the longest range of the measurement area. Then margins ($DiffData$) between the background and the currently scanned data ($CurrData$) are computed for each scanning step (Line 15). Beginning edges of a human object are determined according to two cases: (i) the margin of any scanning step is larger than a threshold TH_{HEB} (Line 17), or (ii) the current data changes more than a threshold TH_{HEC} in one scanning step (Line 20). In our implementation, TH_{HEB} and TH_{HEC} are set to 5 cm and 20 cm, respectively. The current scanning step ($step$) of the detected beginning edges is pushed into a beginning-edge stack (BES_{stack}) for later processing (Lines 18 and 21). End edges are also determined according to two cases: (i) the margin of any scanning steps is smaller than the threshold TH_{HEB} (Line 25), or (ii) the current data changes more than the threshold TH_{HEC} in one scanning step (Line 28). After the system finds the end edge, it calls HUMANEXTRACTION algorithm (Algorithm 2) to extract a human from a group of human objects (Lines 27 and 30).

If the distance between the beginning and end edges is longer than HW_{MIN} and shorter than HW_{MAX} (Line 14), the HUMANEXTRACTION algorithm determines the recognized object as a human object (Line 15). HW_{MIN} and HW_{MAX} are set to 30 cm and 80 cm, respectively.

3.2 Detection of Stationary People

Since we assume moving people are not interested in target objects, we develop STOPDETECTION algorithm (Algorithm 3) to determine whether a human object is moving or stopping. The algorithm finds how long a human move in C cycles. In our implementation, C

Algorithm 1 HUMANDETECTION algorithm.

```
1: procedure HumanDetection
2:   // EndEdge: an end edge of detected object
3:   // BEStack: a stack of beginning edges of objects
4:   // TotalStep: total scanning steps in counter-clockwise direction
5:   // BGData[]: background information
6:   // CurrData[]: currently scanned information
7:   // DiffData[]: the difference between BGData and CurrData
8:   // FlagDetectBE: a flag indicates whether a beginning edge is detected
9:   // THHEB: a threshold for determine human edge when comparing to background information
10:  // THHEC: a threshold for determine human edge when comparing consecutively scanned information
11:  Generate BGData[]
12:  EndEdge ← NULL
13:  BEStack.ToEmpty() // Empty the stack
14:  for step ← 0 to (TotalStep - 1) do
15:    DiffData[step] ← BGData[step] - CurrData[step]
16:    FlagDetectBE ← FALSE
17:    if (BEStack = NULL) & (DiffData[step] > THHEB) then
18:      BEStack.Push(step) // Push step into the stack
19:      FlagDetectBE ← TRUE
20:      else if (BEStack ≠ NULL) & (CurrData[step-1] - CurrData[step] > THHEC) then
21:        BEStack.Push(step)
22:        FlagDetectBE ← TRUE
23:      end if
24:      if (FlagDetectedBE = FALSE) & (BEStack ≠ NULL) then
25:        if DiffData[step] < THHEB then
26:          EndEdge ← step - 1
27:          HUMANEXTRACTION(BEStack, EndEdge)
28:          else if CurrData[step] - CurrData[step - 1] > THHEC then
29:            EndEdge ← step - 1
30:            HUMANEXTRACTION(BEStack, EndEdge)
31:            BEStack.Push(step)
32:          end if
33:        end if
34:        step ← step + 1
35:      end for
36: end procedure
```

is set to five, and the total time of five cycles is 1.065 seconds. First, two candidate positions (*BeginPosition* and *EndPosition*) of a human in C cycle are recorded (Lines 8–16). The difference between two positions, $dist(BeginPosition, EndPosition)$, is the distance that a human covered. If the covered distance is less than or equal to a threshold T_{stop} (30 cm), the person is recognized to be a stationary one (Line 20). Otherwise, the detected object is a moving person (Line 18).

3.3 Detection of Human Direction

Even if people stay near a target object, they are not interested in the target object if their directions do not project onto the object. Therefore, it is necessary to find the direction of a human in order to attain accurate people's attention. The SPAL system uses an angle θ (see Fig. 2) to indicate the direction of a human and then calculate people's attention (Section 4). The angle θ is calculated by DIRECTIONOFHUMAN algorithm (Algorithm 4) as follows.

Algorithm 2 HUMANEXTRACTION algorithm.

```
1: procedure HumanExtraction(BEStack, EndEdge)
2:   // CurrBeginEdge: a current edge to start detecting human
3:   // CurrEndEdge: a current edge to end detecting human
4:   // FlagDetectHumanObject: a flag whether human object is detected
5:   // HWMIN: the minimum width of human
6:   // HWMAX: the maximum width of human
7:   // dist(x, y): a distance between x and y
8:   CurrBeginEdge ← NULL
9:   CurrEndEdge ← EndEdge
10:  FlagDetectHumanObject ← FALSE
11:  while BEStack ≠ NULL do
12:    CurrBeginEdge ← BEStack.Pop()
13:    if FlagDetectHumanObject = FALSE then
14:      if HWMIN < dist(CurrBeginEdge, CurrEndEdge) < HWMAX then
15:        CurrBeginEdge and CurrBeginEdge indicate human
16:        FlagDetectHumanObject ← TRUE
17:      end if
18:      else
19:        CurrBeginEdge and CurrBeginEdge indicate human
20:      end if
21:    end while
22: end procedure
```

Algorithm 3 STOPDETECTION algorithm.

```
1: procedure StopDetection
2:   // C: the number (or cycle) of measurements
3:   // HumanPosition[cycle]: the position of a human at each cycle
4:   // BeginPosition: the beginning position of a person
5:   // EndPosition: the end position of a person
6:   // Tstop: a threshold for determine stopped state
7:   // state: a state of a person (stop or move)
8:   for cycle ← 0 to cyle = C - 1 do
9:     if (HumanPosition[cycle] ≠ NULL) & (BeginPosition = NULL) then
10:      BeginPosition ← HumanPosition[cycle]
11:     end if
12:     if (HumanPosition[C - 1 - cycle] ≠ NULL) & (EndPosition = NULL) then
13:       EndPosition ← HumanPosition[C - 1 - cycle]
14:     end if
15:     cycle ← cycle + 1
16:   end for
17:   if dist(BeginPosition, EndPosition) > Tstop then
18:     state ← move
19:   else
20:     state ← stop
21:   end if
22: end procedure
```

The system computes γ which is the angle of a person with respect to a target object in the polar coordinate system (Line 9). Then it computes a 2D vector to indicate the direction of the person by calling DIRECTIONVECTOR algorithm (Algorithm 5) in Line 10. The angle *LastAngle*, which is an input of the DIRECTIONVECTOR algorithm, is the angle of the vector of a person moving at the last period before changing to a stopping state. Figure 3 illustrates the calculation of *LastAngle*. As the next step, ω which is an absolute angle of the direction of the person in the polar coordinate system is calculated (Line 11). Finally, the system obtains θ , the angle of the

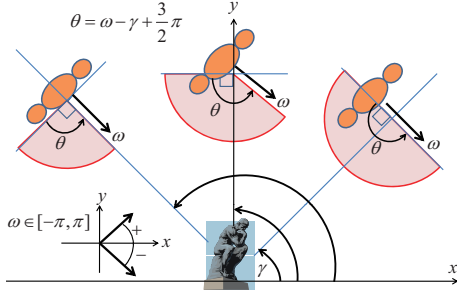


Fig. 2 The angle of human direction (θ) used to calculate attention level.

Algorithm 4 DIRECTIONOFHUMAN algorithm.

```

1: procedure DirectionOfHuman
2:   //  $TotalPerson$ : the number of detected persons
3:   //  $\gamma[\ ]$ : an angle of a person comparing to a target object (see Fig. 2)
4:   //  $Direction[\ ]$ : a 2D vector indicating the direction of a person
5:   //  $LastAngle[\ ]$ : an angle of the last movement vector of a person
6:   //  $\omega[\ ]$ : an absolute angle indicating the direction of a person in the polar coordinate system (see Fig. 2)
7:   //  $\theta[\ ]$ : an angle of the direction of a person relative to a target object (see Fig. 2)
8:   for  $person \leftarrow 0$  to  $TotalPerson - 1$  do
9:      $\gamma[person] \leftarrow \arctan(HumanPosition[person])$ 
10:     $Direction[person] \leftarrow$  DIRECTIONVECTOR( $person, LastAngle[person], \gamma[person]$ )
11:     $\omega[person] \leftarrow \arctan(Direction[person])$ 
12:     $\theta[person] \leftarrow \omega[person] - \gamma[person] + \frac{3}{2}\pi$ 
13:     $person \leftarrow person + 1$ 
14:   end for
15:   return  $\theta[\ ]$ 
16: end procedure

```

human direction relative to a target object (Line 12).

The 2D vector of the direction of a person is calculated by the DIRECTIONVECTOR algorithm (Algorithm 5) as follows. The system computes the middle point ($MidPoint$) between the beginning and end edges (Line 7). Then it finds two normal vectors: a normal vector directed to a target object (V_{obj}) and a normal vector directed to the opposite direction of a target object (V_{opp}) shown in Line 8. If the angle $LastAngle$ is between $\gamma - \frac{3}{2}\pi$ and $\gamma - \frac{1}{2}\pi$, the output is V_{obj} (Line 10). Otherwise, the output is V_{opp} (Line 12).

4. MEASUREMENT MODELS AND METRICS

This section defines two kinds of metrics and proposes two measurement models for calculating the real-world attention which is the last process of the SPAL system (Fig. 1).

4.1 Metrics for Real-world Attention

We define v and V as instantaneous and cumulative values of the real-world attention, respectively. The instantaneous v is the value at a point of time, while the cumulative V is a summation or integration of all v 's in a period of time. The value of attention indicates the num-

Algorithm 5 DIRECTIONVECTOR algorithm.

```

1: procedure DirectionVector( $Person, LastAngle, \gamma$ )
2:   //  $MidPoint$ : the middle point between  $BeginEdge$  and  $EndEdge$ 
3:   //  $V_{obj}$ : a normal vector directed to a target object
4:   //  $V_{opp}$ : a normal vector directed to the opposite direction of a target object
5:    $BeginEdge \leftarrow GetBeginEdge(Person)$ 
6:    $EndEdge \leftarrow GetEndEdge(Person)$ 
7:    $MidPoint \leftarrow FindMidPoint(BeginEdge, EndEdge)$ 
8:    $\{V_{obj}, V_{opp}\} \leftarrow FindVector(BeginEdge, EndEdge)$ 
9:   if  $(\gamma - \frac{3}{2}\pi) < LastAngle < (\gamma - \frac{1}{2}\pi)$  then
10:     $Direction \leftarrow V_{obj}$ 
11:   else
12:     $Direction \leftarrow V_{opp}$ 
13:   end if
14:   return  $Direction$ 
15: end procedure

```

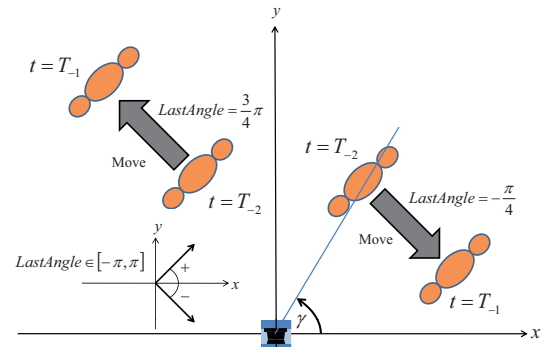


Fig. 3 The angle $LastAngle$ before a person stops walking.

ber of people paying attention to a target object. It is an absolute attention received by each target object. Based on the values of attention, a user can know how many people pay attention to a target object at a point or period of time.

The value of attention highly correlates to the total number of people close to a target object at the time of measurement. It is natural to think that fewer people close to the measurement area leads to a lower value of attention. Therefore, we propose a degree of real-world attention which is a relative attention as the second metric. The degree of attention is a ratio of the number of people paying attention to a target object to the total number of people in a measurement area. Similar to v and V , an instantaneous degree d is the degree of attention at a point of time while a cumulative degree D is the summation or integration of d in a period of time.

4.2 Object-based Attention (OA) Model

An Object-based Attention (OA) model determines the state in which people stay at a certain point for a duration longer than a threshold as a state of paying attention to a target object. We count the number of people (h) who stop nearby an object o as a value of real-world attention. In addition, each counted person ($i = 1, \dots, h$) is weighted by her/his direction towards the target object. Equation (1) expresses the calculation of an instantaneous

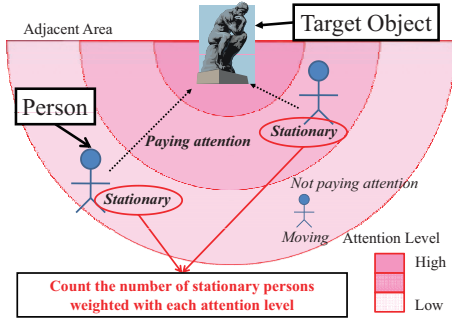


Fig. 4 The distance-weighted attention model.

value of attention (v_{OA}) of the target object o at time t .

$$v_{OA} = \sum_{i=1}^h f(\theta_i). \quad (1)$$

The above equation sums the attention value of all h people which is a function of the angle θ as defined by Eq. (2).

$$f(\theta) = \begin{cases} 0, & \text{if } 0 \leq \theta < \frac{\pi}{4}, \\ \frac{2+\sqrt{2}}{4} \sin(\theta) + \frac{2-\sqrt{2}}{4}, & \text{if } \frac{\pi}{4} \leq \theta \leq \frac{3}{4}\pi, \\ 0, & \text{if } \frac{3}{4}\pi < \theta \leq 2\pi. \end{cases} \quad (2)$$

The value of $f(\theta)$ is between 0.75 and 1 for $\theta \in [\frac{\pi}{4}, \frac{3}{4}\pi]$, otherwise it is zero.

A cumulative value of real-world attention (V) from time t_a to time t_b is an integration of v in the specified period as expressed in Eq. (3).

$$V_{OA} = \int_{t_a}^{t_b} v_{OA} dt. \quad (3)$$

Degrees of real-world attention are calculated by dividing the value of real-world attention by the number of people in a measurement area. Equations (4) and (5) define instantaneous and cumulative degrees of real-world attention, respectively.

$$d_{OA} = \frac{v_{OA}}{h}. \quad (4)$$

$$D_{OA} = \frac{V_{OA}}{\int_{t_a}^{t_b} h dt}. \quad (5)$$

4.3 Distance-weighted Attention (DA) Model

In an actual situation, it is convinced that the shorter the distance between people and a target object, the higher the attention level of people is (Fig. 4). Therefore, a *Distance-weighted Attention (DA) model* includes the distance as a weighting factor (w_i) when calculating the value of real-world attention as expressed in Eq. (6).

$$v_{DA} = \sum_{i=1}^h (f(\theta_i) \cdot w_i). \quad (6)$$

Table 1 The specification of LMS-200 laser scanner.

Laser type	905-nm near infrared
Safety class	1A
Max. scanning angle (°)	180
Angular resolution (°)	0.5
Max. range distance (m)	80
Distance resolution (cm)	1
Scanning rate (Hz)	4.7

The weighting distance is defined in Eq. (7).

$$w_i = \begin{cases} 1, & \text{if } d_i \leq d_l, \\ \frac{d_h - d_i}{d_h - d_l}, & \text{if } d_l < d_i < d_h, \\ 0, & \text{if } d_i \geq d_h, \end{cases} \quad (7)$$

where d_i is the distance between a person i and the target object o at time t ; d_h and d_l ($d_h \geq d_l$) are parameters to control the effect of distance.

An accumulative value of real-world attention can be calculated in the same way as the OA model, i.e., $V_{DA} = \int_{t_a}^{t_b} v_{DA} dt$.

The calculations of instantaneous and accumulative degrees of attention are similar to those of the OA model, i.e., $d_{DA} = (v_{DA})/h$ and $D_{DA} = (V_{DA})/\int_{t_a}^{t_b} h dt$, respectively.

5. SYSTEM IMPLEMENTATION AND EXPERIMENTS

This section presents a prototype implementation of SPAL, followed by the experiment setup and experimental results.

5.1 System Implementation

The SPAL system consists of a laser scanner which connects to a processing node through an RS-232C interface. The laser scanner is LMS-200 developed by SICK in Germany [10]. Table 1 summarizes the specification of the LMS-200 laser scanner. The laser scanner scans in the counter-clockwise direction with a maximum scanning angle of 180 degrees. The angular resolution is 0.5 degree, thereby the laser scanner has 361 scanning steps. The LMS-200 is able to scan at the rate rate of 4.7 Hz. The maximum scanning distance is 80 m and the distance resolution is 1 cm. In the SPAL system, we place the laser scanner 140 cm above the ground level.

The processing node is a laptop computer running Windows Vista, and we use .NET Framework 2.0 as runtime environment. We developed an application software and installed it in the processing node. The application software obtains the scanned raw data from the laser scanner, and then analyzes the data by detecting people and calculating the value and degree of real-world attention. GUI is also developed for easy usage (see Fig. 5). The red trajectory in the figure indicates the front-side circumference of detected objects, i.e., it indicates the dis-

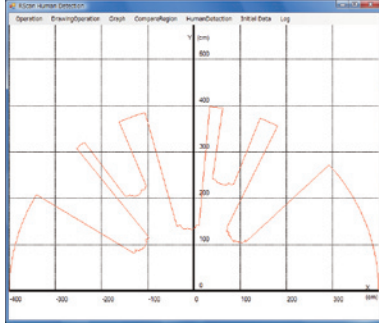


Fig. 5 GUI of application software. The red trajectory indicates the front-side circumference of detected objects.

tance between the laser scanner and detected objects at each scanning step.

5.2 Experiment Setup and Scenarios

The experiments were conducted in a 6 m by 10 m room where a target object and a laser scanner were placed together beside the wall. An area of measurement was set to a half circle with 4-m radius because we assume people outside this area are not interested in the target object. To calculate attention based on the DA model, d_h was set to the longest measurement distance of 4 m, and d_l was set to four values, i.e., 0, 1, 2, and 3 m. We note here that appropriate values of measurement area, d_h , and d_l depend on the type of target object and are adjustable in our system.

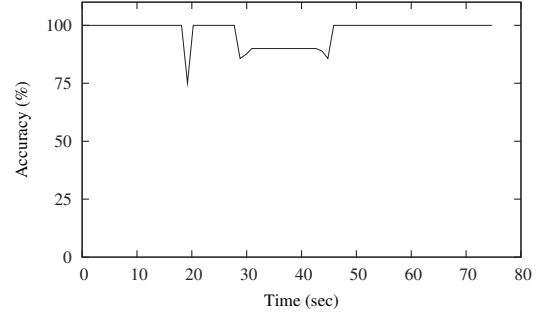
To describe the experimental scenarios and results, we define visitors and noises as follows. The *visitors* are persons who enter the measurement area and pay attention to the target object, while the *noises* indicate persons who enter the measurement area but they are not interested in the object. The noises may walk through the measurement area or they may stop and talk with others in the measurement area. We conducted two experiments as follows.

Experiment I: Three visitors entered the measurement area at 5 s, followed by two and five additional visitors at 20 s and 30 s, respectively. Then five, two, and three visitors left at 45 s, 55 s, and 65 s, respectively.

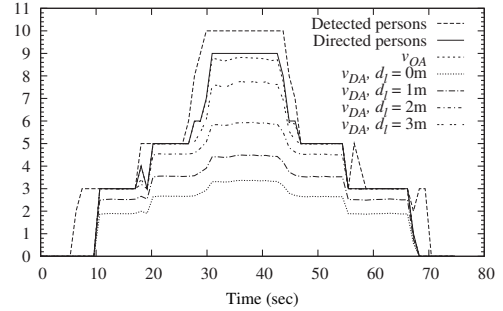
Experiment II: Three noises stayed in the measurement area during the period of the experiment. Three, two, and two visitors entered the area at 23 s, 33 s, and 40 s, respectively. Then two, two, and three visitors left the area at 56 s, 67 s, and 74 s, respectively.

We used a video camera to record the behavior of persons in the measurement area during the experiments. We validated the proposed system based on the video by counting $Attn_a$, the actual number of stationary persons whose directions (θ) point to the target object, i.e., $\frac{\pi}{4} \leq \theta \leq \frac{3\pi}{4}$. The actual value ($Attn_a$) is compared with $Attn_m$, the measured value of stationary persons. The accuracy of the system is then calculated by Eq. (8).

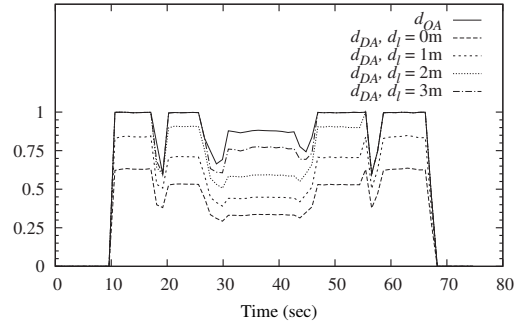
$$\%accuracy = \left(1 - \frac{|Attn_a - Attn_m|}{Attn_a}\right) \times 100 \quad (8)$$



(a) Variation of accuracy along the time.



(b) Instantaneous values of attention (v).

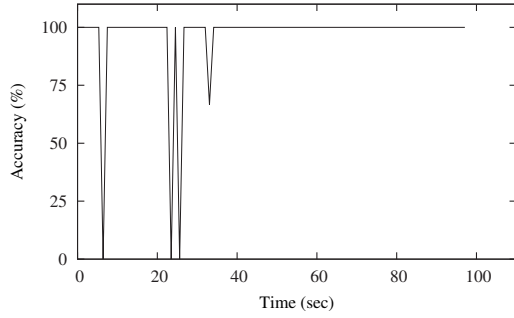


(c) Instantaneous degrees of attention (d).

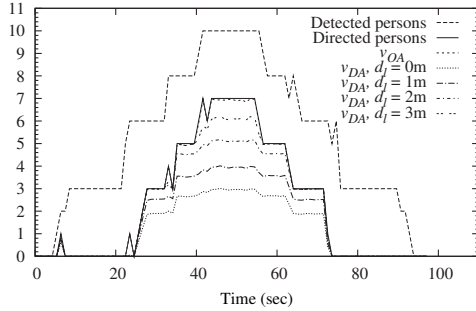
Fig. 6 Experiment I (10 visitors at the maximum).

5.3 Experimental Results

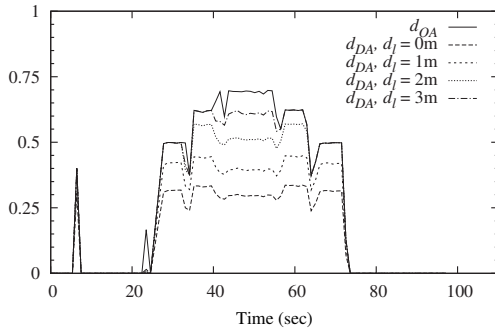
The results of Experiment I are shown in Figs. 6, where “detected persons” means the number of persons detected in the measurement area by our system, and “directed persons” is $Attn_m$ mentioned above. First, we examine the accuracy of $Attn_m$ which is an important value for calculating attention levels. According to Eq. 8, the accuracy of Experiments I is shown in Fig. 6a. At 20 s, the accuracy dropped to 75% because the system takes time to detect and recognize newly visitors. The time required to recognize directed persons depends on the length of measurement cycle (C) used by the STOPDETECTION algorithm (Section 3.2). In addition, the accuracy slightly dropped to 90% when 10 visitors were in the area during 30–45 s, i.e., the system detected only nine out of 10 directed persons. There is no error in other time interval. During 5–10 s, three visitors were walking and their directions did not point toward the target object. Thus the number of detected persons is three (see Fig. 6b), while



(a) Variation of accuracy along the time.



(b) Instantaneous values of attention (v).



(c) Instantaneous degrees of attention (d).

Fig. 7 Experiment II (seven visitors at the maximum with three noises).

$Attn_m$ is still zero. As shown in Fig. 6b, the value of v_{OA} is comparable to $Attn_m$ because v_{OA} is weighted by $f(\theta)$ (Eq. (1)) which is always less than or equal to one. In addition to $f(\theta_i)$, v_{DA} is further attenuated by the weighting distance w_i (Eq. (6)). Therefore, v_{DA} is always lower than or equal to v_{OA} . As expected, the higher the value of d_l is, the higher the value of v_{DA} becomes (Eq. (7)). The degrees of attention are shown in Fig. 6c. The correct value of d_{OA} should be nearly one because all people in the area were visitors. The system shows the correct value of d_{OA} except 30–45 s, where one visitor was not detected correctly. Some steep drops of d are due to changes in the number of directed people in the area. When people enter or leave the area, those moving people are not detected and counted as $Attn_m$, which means the numerator of Eq. (4) decreases while the denominator is still the same. As a result, d always drops when the number of visitors in the area changes.

Table 2 Cumulative values and degrees of attention.

	Experiment I	Experiment II
Period (s)	74.69	97.09
Detected persons	334	473
Directed persons	284	213
V_{OA}	280.07	210.98
V_{DA} ($d_l = 0$ m)	135.29	106.90
V_{DA} ($d_l = 1$ m)	180.39	142.52
V_{DA} ($d_l = 2$ m)	228.77	178.95
V_{DA} ($d_l = 3$ m)	262.09	197.74
D_{OA}	0.839	0.446
D_{DA} ($d_l = 0$ m)	0.405	0.226
D_{DA} ($d_l = 1$ m)	0.540	0.301
D_{DA} ($d_l = 2$ m)	0.685	0.378
D_{DA} ($d_l = 3$ m)	0.785	0.418

Figure 7 shows the results of Experiment II. The number of directed persons is always less than the number of detected persons which means the system is able to extract noises correctly. The accuracy of $Attn_m$ in Fig. 7a is 100% except four step drops due to the same reason as the dropped value at 20 s in Fig. 6a, i.e., the length of measurement cycle (C) used by the STOPDETECTION algorithm. The degrees of attention (Fig. 7c) are nearly the correct values which change from $\frac{3}{6}$, $\frac{5}{8}$, $\frac{7}{10}$, $\frac{5}{8}$, to $\frac{7}{10}$ successively along the time. Note that the numerator is $Attn_m$, and the denominator is the number of people in the measurement area. A larger d_l leads to larger v_{DA} and d_{DA} as the results shown in Fig. 6. We conclude here that the SPAL system is able to measure accurate attention levels even if noises exist in the measurement area.

Table 2 shows cumulative attentions (V and D) of the above experiments. The attentions are accumulated over the periods shown in Figs. 6 and 7. Cumulative attentions show the same trend as instantaneous attentions. In particular, (i) the values of the OA model (V_{OA} and D_{OA}) are approaching the number of directed persons, and (ii) the larger the d_l is, the larger the attention becomes.

The values of attention (v or V) indicate an absolute number of people who pay attention to an object, while the degrees of attention ($0 \leq \{d, D\} \leq 1$) are relative values, i.e., they are the ratio of v or V to the total number of persons in the measurement area. Based on our definitions of attention level, users can know (i) how many people pay attention to an object by using v or V , and (ii) what percentage of people pay attention to an object in the measurement area by using d or D .

The OA model is used when users consider every person in the measurement area has the same weight. Although people are stationary in the area, the model is able to extract those whose directions do not point to a target object. The DA model is applicable in the case where users prefer to weight people according to their distance with respect to a target object. In general, people may stay far away from an object at the first glance, then they move closer to the object if they feel the object is inter-

esting. Therefore, we developed two models for different usages.

6. CONCLUSION AND FUTURE WORK

This paper has studied the tendency of people's attention in the real world by using the proposed SPAL system. The system calculates people's attention based on two proposed measurement models, namely, the object-based and distance-weighted attention models. We have implemented the system and conducted the experiments. Raw data obtained from a laser scanner are input of the system, and people's attention based on two models has been calculated to study the performance of the system. If the calculation time is faster than the period of obtaining raw data, the system is possible to measure people's attention in a real-time fashion. The calculation time required by the experiments in Section 5 is below 1 ms which is much lower than the period of obtaining data (213 ms). Therefore, the proposed algorithms are appropriate for both a real-time measurement and an off-line analysis. The experimental results show that people's attention correlates to the distance between them and a target object.

As one of our future work, we will conduct indoor and outdoor experiments in real environments such as exhibition halls, museums, etc. Finally, we note here a problem of personal privacy. In contrast to existing works on human detection and human tracking, the privacy problem does not arise in the SPAL system because a laser scanner does not recognize an individual.

REFERENCES

- [1] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakeas, and W. von Seelen. Walking pedestrian recognition. *IEEE Transactions on Intelligent Transportation Systems*, 1(3):155–163, Sept. 2000.
- [2] M. Imai, T. Ono, and H. Ishiguro. Physical relation and expression: joint attention for human-robot interaction. *IEEE Transactions on Industrial Electronics*, 50(4):636–643, Aug. 2003.
- [3] N. Ishihara, H. Zhao, and R. Shibasaki. Tracking passenger movement with ground-based laser scanner. In *Proceedings of the Japan Society of Photogrammetry and Remote Sensing Annual Conference*, pages 305–308, June 2002.
- [4] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov. 1998.
- [5] J. H. Lee, Y.-S. Kim, B. K. Kim, K. Ohba, H. Kawata, A. Ohya, and S. Yuta. Security door system using human tracking method with laser range finders. In *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation*, pages 2060–2065, Aug. 2007.
- [6] Y.-F. Ma, X.-S. Hua, L. Lu, and H.-J. Zhang. A generic framework of user attention model and its application in video summarization. *IEEE Transactions on Multimedia*, 7(5):907–919, Oct. 2005.
- [7] Y.-F. Ma, L. Lu, H.-J. Zhang, and M. Li. A user attention model for video summarization. In *Proceedings of the 10th ACM International Conference on Multimedia*, pages 533–542, Dec. 2002.
- [8] Y.-F. Ma and H.-J. Zhang. A model of motion attention for video skimming. In *Proceedings of the IEEE International Conference on Image Processing*, pages 129–132, Sept. 2002.
- [9] K. Nakamura, H. Zhao, R. Shibasaki, K. Sakamoto, and N. Suzukawa. Visualizing passenger flow in railway station using laser scanners. In *Proceedings of the 9th International Conference on Computers in Urban Planning and Urban Management*, 2005.
- [10] SICK. <http://www.sick.com>.
- [11] C. Wöhler, J. K. Anlauf, T. Pörtner, and U. Franke. A time delay neural network algorithm for real-time pedestrian recognition. In *Proceedings of the IEEE International Conference on Intelligent Vehicles*, pages 247–252, 1998.
- [12] H. Zhao, K. Nakamura, and R. Shibasaki. A laser-scanner system for acquiring walking trajectory data and its possible application to behavioral science. In A. Okabe, editor, *GIS-based Studies in the Humanities and Social Sciences*, chapter 3. Taylor & Francis/CRC Press, Nov. 2005.
- [13] H. Zhao, X. Shao, K. Katabira, and R. Shibasaki. Joint tracking and classification of moving objects in cross road using single-row laser range scanner. In *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, pages 287–294, Sept. 2006.
- [14] H. Zhao and R. Shibasaki. Pedestrian tracking using multiple laser range scanners. In *Proceedings of the International Conference on Computers in Urban Planning and Urban Management*, 2003.
- [15] L. Zhao and C. E. Thorpe. Stereo- and neural network-based pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 1(3):148–154, Sept. 2000.