IPSJ Transactions on Advanced Computing Systems Vol. 2 No. 1 173–186 (Mar. 2009)

Regular Paper

Observing Real-world Attention by a Laser Scanner

Yasunori Yakiyama,†1 Niwat Thepvilojanapong,†1,†2
Masayuki Iwai,†1,†2 Oru Mihirogi,†1
Kazunori Umeda†2,†3 and Yoshito Tobe†1,†2

Although human activities in the World Wide Web are increasing rapidly 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 SPAL, a Sensor of Physical-world Attention using Laser scanning. 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 to 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.

1. Introduction

Laser scanners and image sensors have been used for measuring human behavior [12]. 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),14)]. 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. For example, we can place objects at the right places in an exhibition hall or a store according to people's attention.

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. People's attention can be analyzed from images captured by a stereo camera. 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 overlapped humans. To identify whether or not 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 addition, a human highly interested will stop to watch a target object, we propose an algorithm to determine stationary people from all detected humans. After establishing people who pay attention to a target object, the system calculates real-world attention levels according to two proposed models: *Object-based Attention (OA)* and *Distance-weighted Attention (DA)* models. We 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 SPAL system including human detection and measurement models. The experimental results show high accuracy of 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 define two metrics to evaluate people's attention.
- We propose two measurement models to compute people's attention.
- We conducted experiments to evaluate our proposed system.

The rest of this paper is organized as follows. Section 2 discusses related work.

†1 Tokyo Denki University
†2 CREST, JST
†3 Chuo University

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 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 scope 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 and multiple laser scanners and analyzed walking trajectory based on a pedestrian model[3,12,13]. 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 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 also used to recognize pedestrians from images[11,15]. Previous works use laser scanners and stereo cameras as two major devices for tracking and detecting the flow of people. Although a stereo camera can be used to measure a distance from an object, it is affected by some environment parameters such as the light and poses a privacy problem in public spaces. Consequently, we consider using laser scanners in our work.

3. Human Detection System

The overall processing flow of the SPAL system is shown in **Fig. 1**. This section presents the algorithms used by the SPAL system to detect people who pay attention to a target object. A process of human detection consists of three main steps. First, the system detects an individual human when people come into a measurement area (Section 3.1). Second, the system determines whether a person is moving or stationary at a certain point (Section 3.2). Third, the system detects the direction of a human to determine whether the detected person is watching a target object or not (Section 3.3).

3.1 Detection of Human Object

The system performs a step of detecting human objects by using the HumanDetection algorithm (**Fig. 2**). The algorithm is able to extract a human from a group of overlapped human objects. We assume a laser scanner scans in the counterclockwise direction when briefly describing the algorithm as follows. 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). If any objects appear in the measurement area, scanned data are always less than the background data. Thus the *DiffData* are non-negative values. 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}

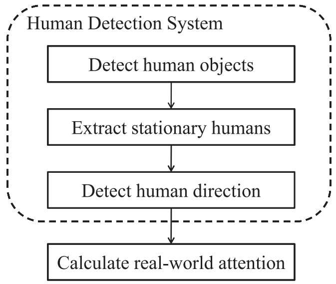


Fig. 1 The processing flows of SPAL system.

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: // TH_{HEB}: a threshold for determine human edge when comparing to background information
10: // TH_{HEC}: a threshold for determine human edge when comparing consecutively scanned information
11: Generate $BGData[\,]$
12: $EndEdge \leftarrow$ NULL
13: $BEStack.ToEmpty()$ // Empty the stack
14: **for** $step \leftarrow 0$ **to** $(TotalStep - 1)$ **do**
15: $DiffData[step] \leftarrow BGData[step] - CurrData[step]$
16: $FlagDetectBE \leftarrow$ FALSE
17: **if** $(BEStack =$ NULL$)$ & $(DiffData[step] > TH_{HEB})$ **then**
18: $BEStack.Push(step)$ // Push $step$ into the stack
19: $FlagDetectBE \leftarrow$ TRUE
20: **else if** $(BEStack \neq$ NULL$)$ & $(CurrData[step-1] - CurrData[step] > TH_{HEC})$ **then**
21: $BEStack.Push(step)$
22: $FlagDetectBE \leftarrow$ TRUE
23: **end if**
24: **if** $(FlagDetectedBE =$ FALSE$)$ & $(BEStack \neq$ NULL$)$ **then**
25: **if** $DiffData[step] < TH_{HEB}$ **then**
26: $EndEdge \leftarrow step - 1$
27: HUMANEXTRACTION$(BEStack, EndEdge)$
28: **else if** $CurrData[step] - CurrData[step-1] > TH_{HEC}$ **then**
29: $EndEdge \leftarrow step - 1$
30: HUMANEXTRACTION$(BEStack, EndEdge)$
31: $BEStack.Push(step)$
32: **end if**
33: **end if**
34: $step \leftarrow step + 1$
35: **end for**
36: **end procedure**

Fig. 2 HUMANDETECTION algorithm.

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: // HW_{MIN}: the minimum width of human
6: // HW_{MAX}: the maximum width of human
7: // $dist(x, y)$: a distance between x and y
8: $CurrBeginEdge \leftarrow$ NULL
9: $CurrEndEdge \leftarrow EndEdge$
10: $FlagDetectHumanObject \leftarrow$ FALSE
11: **while** $BEStack \neq$ NULL **do**
12: $CurrBeginEdge \leftarrow BEStack.Pop()$
13: **if** $FlagDetectHumanObject =$ FALSE **then**
14: **if** $HW_{MIN} < dist(CurrBeginEdge, CurrEndEdge) < HW_{MAX}$ **then**
15: $CurrBeginEdge$ and $CurrBeginEdge$ indicate human
16: $FlagDetectHumanObject \leftarrow$ TRUE
17: **end if**
18: **else**
19: $CurrBeginEdge$ and $CurrBeginEdge$ indicate human
20: **end if**
21: **end while**
22: **end procedure**

Fig. 3 HUMANEXTRACTION algorithm.

(Line 17), (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 and 20 cm, respectively. The current scanning step (*step*) of the detected beginning edges is pushed into a beginning-edge stack (*BEStack*) for later processing (Line 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), (ii) current data changes more than the threshold TH_{HEC} in one scanning step (Line 28). After the system finds the end edge, it calls the HUMANEXTRACTION algorithm (**Fig. 3**) to extract a human object (Line 27 and 30). In short, if the distance between the beginning and end edges is longer than HW_{MIN} and shorter than HW_{MAX} (Line 14), the recognized object is determined as a human object (Line 15). HW_{MIN} and HW_{MAX} are set to 30 and 80 cm, respectively.

The proposed algorithm will detect a human correctly, if the edges of each person are separated by at least 1°. In particular, each person should separate at least two times the angular resolution of a laser scanner which is equal to 0.5° in our implementation. The distance (in cm) between two persons, which depends on the distance from the scanner, is $\delta\frac{\pi}{180}$, where δ is an angle (in degrees) between the edges of each person.

We note that our system assumes overlapped people do not pay attention to a target object. The reason is that overlapped people are those who stay behind someone and they cannot watch a target object. In some cases, they may be able to watch a target object with some difficulty due to obstruction by the people in front of them, consequently they have very little attention. If people are really interested in a target object, they will not stay behind someone, i.e., they will change their positions from the obstruction in order to watch the target object clearly.

3.2 Detection of Stationary People

We assume moving people are not interested in target objects. Therefore we develop a StopDetection algorithm (**Fig. 4**) to determine whether a human object is moving or stopping. The system finds how long a human move in C cycles. In our implementation, C is set to five and the total time of five cycles is 1.065 seconds. First, two candidate positions (*BeginPosision* and *EndPosition*) of a human in C cycle are selected (Line 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 17–21).

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 determine people's attention correctly. The SPAL system uses an angle θ (cf. **Fig. 5**) to indicate the direction of a human and then calculate people's attention. The angle θ is calculated by the DirectionOfHuman algorithm (**Fig. 6**) as follows.

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 the DirectionVector algorithm (**Fig. 7**). 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 8** illustrates the calculation of *LastAngle*.

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:      // T_stop: a threshold for determine stopped state
7:      // state: a state of a person (stop or move)
8:      for cycle ← 0 to 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) > T_stop then
18:         state ← move
19:     else
20:         state ← stop
21:     end if
22: end procedure
```

Fig. 4 StopDetection algorithm.

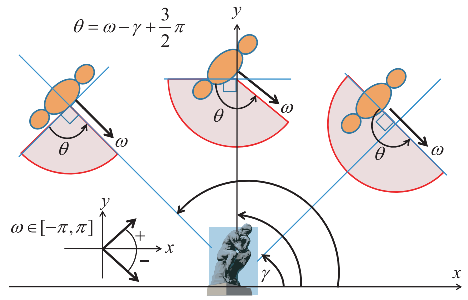


Fig. 5 The angle of human direction (θ).

Algorithm 4 DirectionOfHuman algorithm.

```
1:  procedure DirectionOfHuman
2:      // TotalPerson: the number of detected persons
3:      // γ[ ]: an angle of a person comparing to a target object (cf. Fig. 5)
4:      // Direction[ ]: a 2D vector indicating the direction of a person
5:      // LastAngle[ ]: an angle of the last movement vector of a person
6:      // ω[ ]: an absolute angle indicating the direction of a person in the polar coordinate system (cf. Fig. 5)
7:      // θ[ ]: an angle of the direction of a person relative to a target object (cf. Fig. 5)
8:      for person ← 0 to TotalPerson − 1 do
9:          γ[person] ← arctan(HumanPosition[person])
10:         Direction[person] ← DirectionVector(person, LastAngle[person], γ[person])
11:         ω[person] ← arctan(Direction[person])
12:         θ[person] ← ω[person] − γ[person] + 3/2 π
13:         person ← person + 1
14:     end for
15:     return θ[ ]
16: end procedure
```

Fig. 6 DirectionOfHuman algorithm.

Algorithm 5 DirectionVector algorithm.

```
1:  procedure DirectionVector(Person, LastAngle, γ)
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 ← GetBeginEdge(Person)
6:      EndEdge ← GetEndEdge(Person)
7:      MidPoint ← FindMidPoint(BeginEdge, EndEdge)
8:      {V_obj, V_opp} ← FindVector(BeginEdge, EndEdge)
9:      if (γ − 3/2 π) < LastAngle < (γ − π/2) then
10:         Direction ← V_obj
11:     else
12:         Direction ← V_opp
13:     end if
14:     return Direction
15: end procedure
```

Fig. 7 DirectionVector algorithm.

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

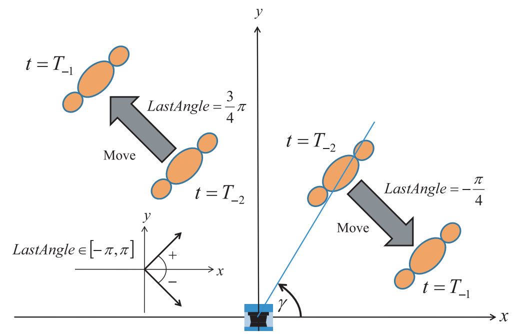


Fig. 8 The angle *LastAngle* before a person stops walking.

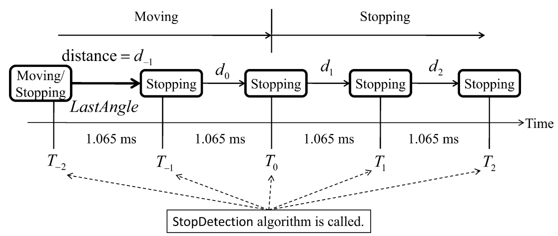


Fig. 9 Time sequence when a person changes to a stopping state.

The 2D vector of the direction of a person is calculated by the DirectionVector algorithm (Fig. 7) 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 the vector directed to a target object (V_{obj}). Otherwise, the output is the vector directed to the opposite direction of a target object (V_{opp}).

To detect stationary people and directions correctly, there are some limitations on the tolerated distance and change in the direction when a person stops. **Figure 9** shows the time sequence when a person changes to a stopping state. The StopDetection algorithm is called and processed every C cycles (1.065 ms), i.e., at time T_{-2}, T_{-1}, T_0, T_1, and T_2 in the figure. d_k is the distance that a person has covered from time T_{k-1} to time T_k. We note again that *LastAngle* is the angle of the vector of the person moving from time T_{-2} to time T_{-1} before changing to a stopping state.

There are two quantitative conditions to detect stationary people and directions correctly.

(1) When a person changed to a stopping state at time T_0, the person may not change its direction more than $\pm 180°$, and may not move more than 30 cm. These conditions can be expressed as the following equations.

$$\left(\gamma - \frac{3}{2}\pi\right) < LastAngle < \left(\gamma - \frac{\pi}{2}\right) \tag{1}$$

$$d_0 \leq T_{stop} \tag{2}$$

(2) After changing to a stopping state, a person may not move more than 30 cm until he changes to a moving state as expressed by Eq. (3).

$$d_t \leq T_{stop}, \quad \text{where } t = 1, 2, \ldots \tag{3}$$

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 of v in a specified interval. The value of attention indicates the number of people paying attention to a target object. It is an absolute attention paid to each target object. A user can know how many people pay attention to the target object at a point of time or in a specified interval.

The value of attention highly correlates the total number of people close to a measurement area. 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 between the number of people paying attention to a target object and the 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 of d in a specified interval. We note here that both the value and degree of attention are useful in different situations.

4.2 Object-based Attention 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 value as a state of paying attention to a target object. We count the number of people (N_h) who stop nearby an object o as a value of real-world attention. In addition, each counted person ($h = 1, \ldots, N_h$) is weighted by his direction towards the target object. Equation (4) expresses the calculation of an instantaneous value of attention (v_{OA}) of the target object o at time t.

$$v_{OA}(o,t) = \sum_{h=1}^{N_h(t)} s(o,h,t). \tag{4}$$

The value is summed for all people N_h and weighted by the function $s(o,h,t)$ whose value is determined by the angle θ.

$$s(o,h,t) = f(\theta(o,h,t)). \tag{5}$$

The angle θ is measured in the counterclockwise direction from the plane perpendicular to the target object (Fig. 5). The function for calculating the direction-based weight is expressed in Eq. (6).

$$s(o,h,t) = \begin{cases} 0, \quad 0 \leq \theta < \frac{\pi}{4} \\ \frac{2+\sqrt{2}}{4}\sin(\theta) + \frac{2-\sqrt{2}}{4}, \quad \frac{\pi}{4} \leq \theta \leq \frac{3}{4}\pi \\ 0, \quad \frac{3}{4}\pi < \theta \leq 2\pi. \end{cases} \tag{6}$$

The weight 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 $v(t)$ in the specified period as shown in Eq. (7).

$$V_{OA}(o,t_a,t_b) = \int_{t_a}^{t_b} v_{OA}(o,t)dt. \tag{7}$$

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 (8) and (9) express instantaneous and cumulative degree of real-world attention, respectively.

$$d_{OA}(o,t) = \frac{v_{OA}(o,t)}{N_h(o,t)}. \tag{8}$$

$$D_{OA}(o,t_a,t_b) = \frac{V_{OA}(o,t_a,t_b)}{\int_{t_a}^{t_b} N_h(o,t)dt}. \tag{9}$$

4.3 Distance-weighted Attention Model

In an actual situation, it is considered that the shorter the distance between people and a target object, the higher the attention level of people is (**Fig. 10**). Therefore, a *Distance-weighted Attention (DA) model* includes the distance as a weighting factor ($w(o,h,t)$) when calculating the value of real-world attention (Eq. (10)).

$$v_{DA}(o,t) = \sum_{h=1}^{N_h(o,t)} \{s(o,h,t) \cdot w(o,h,t)\}. \tag{10}$$

The weighting distance is defined in Eq. (11).

$$w(o,h,t) = \begin{cases} 1, & dist(o,h,t) \leq d_1 \\ \dfrac{d_0 - dist(o,h,t)}{d_0 - d_1}, & d_1 < dist(o,h,t) < d_0 \\ 0, & d_0 \leq dist(o,h,t). \end{cases} \tag{11}$$

$dist(o,h,t)$ is the distance between a person h and the target object o at time t.

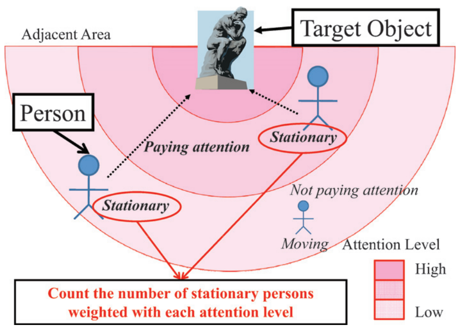


Fig. 10 The distance-weighted attention model.

d_0 and d_1 are some parameters to control the effect of distance, where $d_0 \geq d_1$.

An accumulative value of real-world attention can be calculated in the same way as the OA model (Eq. (7)) by integrating $v_{DA}(o,t)$ from t_a to t_b.

Instantaneous and accumulative degrees of real-world attention are calculated in the same way as the OA model (Eqs. (8) and (9)).

For both OA and DA models, if desired, the average value of real-world attention (V) can be calculated by using Eq. (12).

$$\overline{V_{\{OA,DA\}}(o,t_a,t_b)} = \frac{V_{\{OA,DA\}}(o,t_a,t_b)}{t_b - t_a}. \tag{12}$$

5. Implementation and Experiments

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

5.1 Prototype 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** shows 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 degrees which means the laser scanner has 361 scanning steps. 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. Consequently, the scanning plane is approximately at the level of the chest of an adult. The scanning rate is 4.7 Hz.

An illustration of the scanning process is represented in **Fig. 11**. The laser scanner measures the distances between the laser scanner itself and detected objects. As stated above, there are 361 scanning steps from 0 to 180 degrees

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

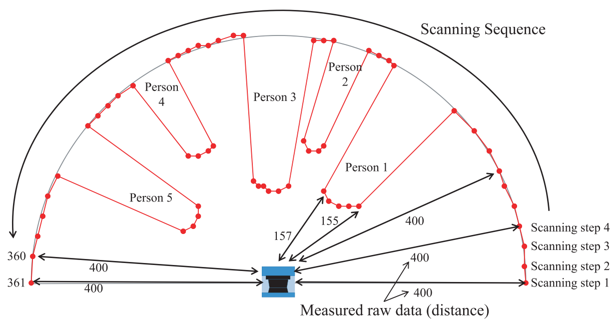


Fig. 11 Scanning sequence and measured raw data.

with an angular resolution of 0.5 degrees. The maximum scanning distance in the experiments is set to 400 cm, consequently if an object does not exist in any scanning step, the value of the raw data is 400. The measured distances (in centimeter) in an increasing order of scanning steps are shown in **Fig. 12** from left to right and top to bottom, respectively. Each row in the figure shows the measured value (or raw data) of ten scanning steps. The leftmost column shows the order of scanning steps where x is substituted with the values in the topmost row. According to the algorithms described above, the SPAL system is able to detect five persons from the measured raw data shown in Fig. 12.

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 (**Fig. 13**) is also developed for easy usage. The red trajectory in the figure indicates the front-side circumference of detected objects. The front-side circumference is the value of measured raw data, i.e., the distance between the laser scanner and detected objects. The trajectory illustrated in Fig. 13 corresponds to the measured raw data shown in Fig. 12.

As described in Sections 3 and 4, our system is sensitive to the angle of observed people which is used to identify the value and degree of attention. We performed preliminary experiments to investigate the accuracy of γ which is a raw data

x =	1	2	3	4	5	6	7	8	9	10
x	400	400	400	400	400	400	400	400	400	400
1x	400	400	400	400	400	400	400	400	400	400
2x	400	400	400	400	400	400	400	400	400	400
3x	400	400	400	400	400	400	400	400	400	400
4x	400	400	400	400	400	400	400	400	400	400
5x	400	400	400	400	400	400	400	400	400	400
6x	400	400	400	400	400	400	400	400	400	400
7x	400	400	400	400	400	400	400	400	400	400
8x	400	400	400	400	400	400	400	155	155	156
9x	147	147	147	147	145	143	144	143	143	142
10x	142	141	140	138	139	139	139	139	139	140
11x	140	140	142	143	142	143	142	145	148	148
12x	147	147	147	149	157	157	400	400	400	400
13x	400	400	400	400	400	400	400	400	400	247
14x	245	244	244	245	245	241	239	240	238	239
15x	239	239	239	240	240	239	241	244	245	245
16x	246	246	400	400	400	400	400	400	400	400
17x	400	142	143	143	144	144	144	138	138	137
18x	137	136	134	133	133	133	134	135	135	135
19x	135	135	136	136	136	136	136	136	135	136
20x	137	139	140	142	147	147	146	146	145	146
21x	145	400	400	400	400	400	400	400	400	400
22x	400	400	400	400	400	400	400	400	400	249
23x	249	246	245	244	246	246	244	242	242	241
24x	241	241	241	241	242	244	246	247	252	250
25x	250	252	254	400	400	400	400	400	400	400
26x	156	152	151	150	151	152	153	153	148	149
27x	148	146	145	145	145	145	144	144	144	144
28x	145	144	145	145	146	148	148	147	149	153
29x	156	154	153	153	153	153	155	400	400	400
30x	400	400	400	400	400	400	400	400	400	400
31x	400	400	400	400	400	400	400	400	400	400
32x	400	400	400	400	400	400	400	400	400	400
33x	400	400	400	400	400	400	400	400	400	400
34x	400	400	400	400	400	400	400	400	400	400
35x	400	400	400	400	400	400	400	400	400	400
36x	400									

Fig. 12 An example of raw data obtained from the laser scanner. The SPAL system analyzes the raw data and creates the corresponding GUI as illustrated in Fig. 13.

input from the laser scanner and is used for other calculations including people's attention. The errors of γ are shown in **Table 2** where the measured values are averaged over ten experiments. The results show that the error of the observed angle is within $\pm 2°$. According to the proposed metrics and models for real-world

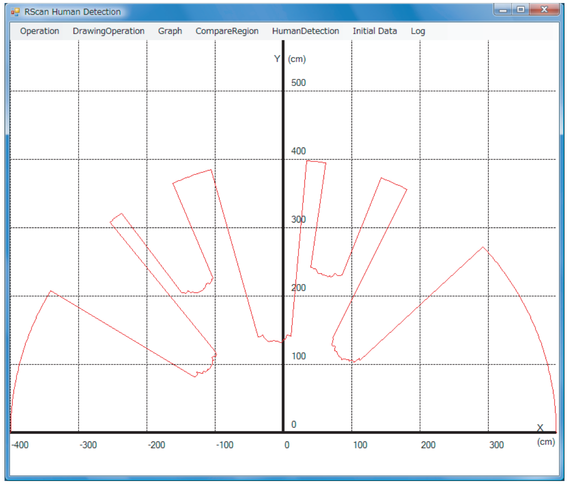


Fig. 13 GUI of application software. The trajectory indicates the front-side circumference of detected objects.

Table 2 Accuracy of γ.

γ	Average	Error
45°	47.47°	2.47°
90°	89.89°	−0.11°
135°	132.40°	−2.60°

attention, the error of $\pm 2^\circ$ affects 2% of the calculated attention.

5.2 Experiment Setup

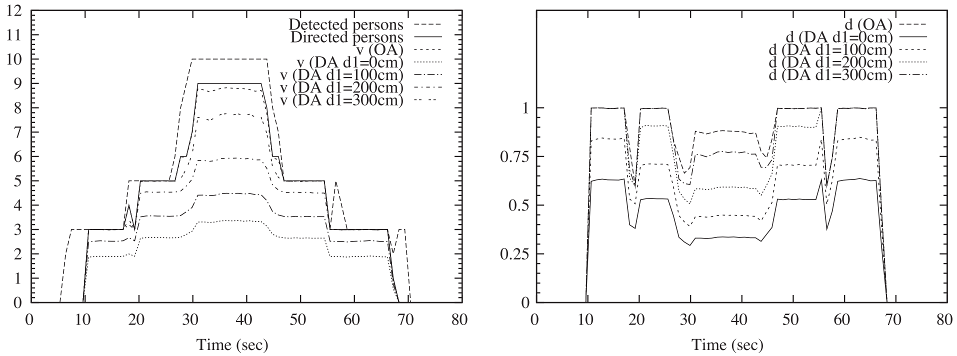
The experiments were conducted in a 6 m by 10 m room where a target object and a laser scanner were placed together beside the wall. An area of human detection and measurement was set to a half circle with 4-m radius because we assume people outside this area are not interested in the target object. Note that the measurement area is adjustable and depends on the types of target objects. For the DA model, d_0 is set to the longest measurement distance (4 m), and d_1 is set to 0, 1, 2, and 3 meters. Note that the appropriate d_1 depends on the types of target objects.

To describe the experiments 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 mean persons who enter the measurement area but they are not interested in the object, i.e., they may walk through the measurement area or they may stop and talk with others in the measurement area. The number of visitors and noises varied along the time. We used a video recorder to record the behavior of persons in the measurement area. 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., $\frac{\pi}{4} \leq \theta \leq \frac{3}{4}\pi$. We compare the actual value ($Attn_a$) with $Attn_m$, the measured number of stationary persons whose directions (θ) point to the target object, where θ is obtained from Algorithm 4 (Fig. 6). The accuracy of the system is then calculated by Eq. (13).

$$\%accuracy = \left(1 - \frac{|Attn_a - Attn_m|}{Attn_a}\right) \times 100 \qquad (13)$$

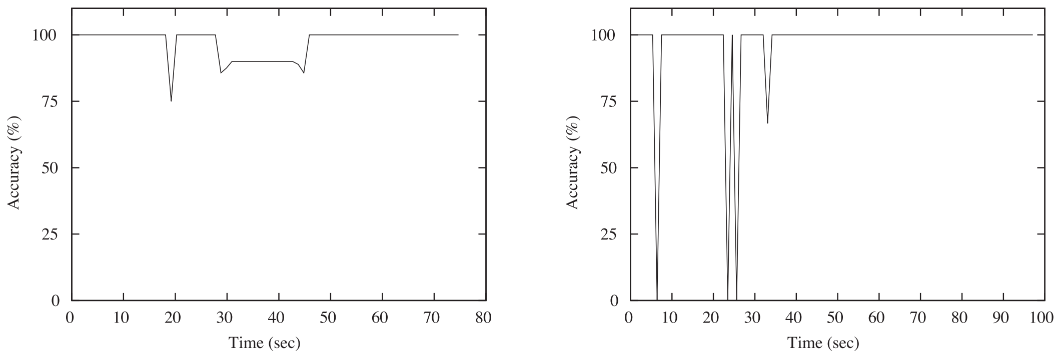
5.3 Experimental Results

Although the experiments were conducted for a long period of time, we show some parts of experiments when closely investigating the results. **Fig. 14** (a) shows the results of a situation where three visitors entered the measurement area at 5 s, followed by two and five additional visitors at 20 and 30 s, respectively. Then five, two, and three visitors left at 45, 55, and 65 s, respectively. In the figure, "Detected persons" means the number of persons detected in the measurement area by our system, and "Directed persons" is $Attn_m$ mentioned above. As a reference, the raw data obtained from the laser scanner at an instance of time (52.079 s) are shown in Fig. 12, and the corresponding trajectory is illustrated in Fig. 13. First, we examine the accuracy of $Attn_m$ which is an important value for calculating attention levels. The accuracy of Fig. 14 (a) is shown in **Fig. 15** (a). At 19 s, the accuracy dropped to 75% because the system requires time to detect directed persons. The required time depends on the cycle of measurement (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 28–45 s, i.e., the system detected only nine out of 10 directed persons. There is no error in any 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 while $Attn_m$ is zero. The value of v_{OA} is



(a) Instantaneous values of attention (v). (b) Instantaneous degrees of attention (d).

Fig. 14 Visitors vary from 0–10 persons along the time.

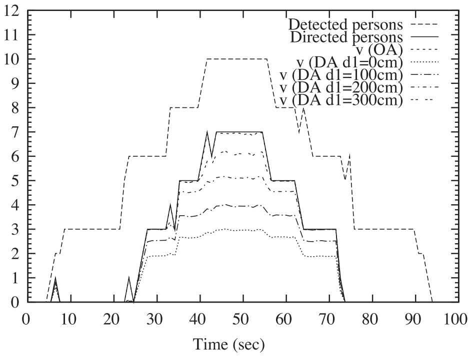


(a) 10 visitors at the maximum (Fig. 14). (b) 7 visitors at the maximum with three noises (Fig. 16).

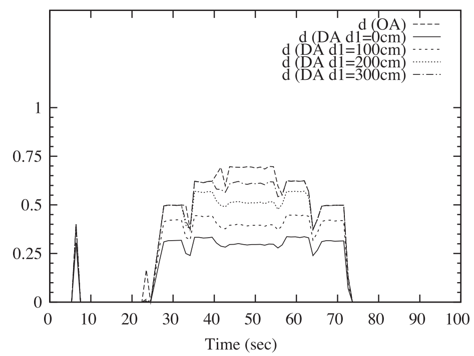
Fig. 15 Accuracy of all time intervals (Figs. 14 and 16).

comparable to $Attn_m$ because v_{OA} is weighted by $s(o, h, t)$ (Eq. (4)) which is always less than or equal to one. In addition to $s(o, h, t)$, v_{DA} is further attenuated by the weighting distance (Eq. (10)). Therefore, v_{OA} is lower than v_{DA}. As expected, the higher the value of d_1 is, the higher the value of v_{DA} becomes (Eq. (11)). The degrees of attention are shown in Fig. 14 (b). The correct value of d_{OA} should be nearly one because all people in the area are 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. (8) decreases while the denominator is still the same. As a result, d always drops when the number of visitors in the area changes.

Figure 16 shows the results of a situation where three noises stayed in the area all the time, as three, two, and two visitors entered the area, respectively. The



(a) Instantaneous values of attention (v).



(b) Instantaneous degrees of attention (d).

Fig. 16 Visitors vary from 0–7 persons along the time. Three noises stay in the area all the times.

Table 3 Cumulative values of attention (the values in parenthesis under V_{DA} is d_1).

	Period (s)	Detected persons	Directed persons	V_{OA}	V_{DA} (0 cm)	V_{DA} (100 cm)	V_{DA} (200 cm)	V_{DA} (300 cm)
10 visitors at the max.	74.69	334	284	280.07	135.29	180.39	228.77	262.09
7 visitors at the max. with three noises	97.09	473	213	210.98	106.90	142.52	178.95	197.74

Table 4 Cumulative degrees of attention (the values in parenthesis under D_{DA} is d_1).

	Period (s)	Detected persons	Directed persons	D_{OA}	D_{DA} (0 cm)	D_{DA} (100 cm)	D_{DA} (200 cm)	D_{DA} (300 cm)
10 visitors at the max.	74.69	334	284	0.839	0.405	0.540	0.685	0.785
7 visitors at the max. with three noises	97.09	473	213	0.446	0.226	0.301	0.378	0.418

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. 15 (b) is 100% except four steep drops due to the same reason as the dropped value at 19 s in Fig. 15 (a), i.e., the cycle of measurement (C) used by the StopDetection algorithm. The degrees of attention (Fig. 16 (b)) 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_1 leads to larger v_{DA} and d_{DA} as the results shown in Fig. 14.

Tables 3 and **4** show cumulative attentions (V and D) of the above experiments. The attentions are accumulated over the interval shown in Fig. 14 (10 visitors at the maximum) and Fig. 16 (seven visitors at the maximum with three noises). 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 d_1 is, the larger the attention becomes.

The value of attention (v or V) indicates an absolute number of people who pay attention to an object, while the degree of attention ($0 \leq \{d, D\} \leq 1$) is a relative value, i.e., the value v or V with respect to the number of persons in the area. Users 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 comparison with other objects 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 an object. In general, people may stay far away from an object at the first glance, they then move closer to the object if they feel the object is interesting. 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 OA and DA models. We have implemented the system and conducted the experiments. Raw data obtained from laser scanners are inputs of the system, and people's attention based on three 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 the distance between them and a target object.

If the calculation time is longer than the period of obtaining raw data, we recommend to apply the following solutions in order to avoid accuracy fall: (i) reduce the scanning rate of the laser scanner, (ii) increase the value of C used in the StopDetection algorithm, and (iii) use a buffer large enough to keep raw data before calculation.

As one of our future work, we will conduct indoor and outdoor experiments in real environments such as events, 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) Curio, C., Edelbrunner, J., Kalinke, T., Tzomakeas, C. and von Seelen, W.: Walking pedestrian recognition, *IEEE Trans. Intelligent Transportation Systems*, Vol.1, No.3, pp.155–163 (2000).
2) Imai, M., Ono, T. and Ishiguro, H.: Physical relation and expression: Joint attention for human-robot interaction, *IEEE Trans. Industrial Electronics*, Vol.50, No.4, pp.636–643 (2003).
3) Ishihara, N., Zhao, H. and Shibasaki, R.: Tracking passenger movement with ground-based laser scanner, *Proc. Japan Society of Photogrammetry and Remote Sensing (JSPRS) Annual Conference*, pp.305–308 (2002).
4) Itti, L., Koch, C. and Niebur, E.: A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Analysis Machine Intelligent*, Vol.20, No.11, pp.1254–1259 (1998).
5) Lee, J.H., Kim, Y.-S., Kim, B.K., Ohba, K., Kawata, H., Ohya, A. and Yuta, S.: Security Door System Using Human Tracking Method with Laser Range Finders, *Proc. 2007 IEEE International Conference on Mechatronics and Automation*, pp.2060–2065 (2007).
6) Ma, Y.-F., Hua, X.-S., Lu, L. and Zhang, H.-J.: A generic framework of user attention model and its application in video summarization, *IEEE Trans. Multimedia*, Vol.7, No.5, pp.907–919 (2005).
7) Ma, Y.-F., Lu, L., Zhang, H.-J. and Li, M.: A user attention model for video summarization, *Proc. 10th ACM International Conference on Multimedia (MULTIMEDIA'02)*, Juan-les-Pins, France, pp.533–542 (2002).
8) Ma, Y.-F. and Zhang, H.-J.: A model of motion attention for video skimming, *Proc. IEEE International Conference on Image Processing (ICIP'02)*, pp.129–132 (2002).
9) Nakamura, K., Zhao, H., Shibasaki, R., Sakamoto, K. and Suzukawa, N.: Visualizing passenger flow in railway station using laser scanners, *Proc. 9th International Conference on Computers in Urban Planning and Urban Management* (2005).
10) SICK: Laser Measurement Systems. available from http://www.sick.com/home/factory/catalogues/auto/lmsindoor/en.html
11) Wohler, J.C., Anlauf, J., Portner, T. and Franke, U.: A time delay neural network

algorithm for real-time pedestrian recognition, *Proc. Intelligent Vehicles Symposium (IV'98)*, pp.247–252 (1998).
12) Zhao, H., Nakamura, K. and Shibasaki, R.: A laser scanner system for acquiring walking trajectory data and its possible applications to behavioral science, *Studies in Human and Social Sciences with GIS (Chapter 3)*, 084932713x, Taylor & Francis/CRC Press (2005).
13) Zhao, H. and Shibasaki, R.: Pedestrian Tracking using Multiple Laser Range Scanners, *Proc. 8th International Conference on Computers in Urban Planning and Urban Management* (2003).
14) Zhao, H., Shao, X., Katabira, K. and Shibasaki, R.: Joint tracking and classification of moving objects in cross road using single-row laser range scanner, *Proc. IEEE International Conference on Intelligent Transportation Systems*, pp.287–294 (2006).
15) Zhao, L. and Thorpe, C.: Stereo and neural network-based pedestrian detection, *Proc. IEEE International Conference on Intelligent Transportation Systems*, pp.289–303 (1999).

(Received July 23, 2008)
(Accepted January 16, 2009)

Yasunori Yakiyama was born in 1978. He received his M.E. from Kyushu University in 2002, and was engaged in research on communication systems with Toshiba Corporation from 2002. He became a researcher at Tokyo Denki University in 2005. He has been Ph.D. student at Tokyo Denki University since 2006. His research interests include multimedia communications, sensor networks, and ubiquitous computing. He is a member of IEEE Computer Society, ACM and IPSJ.

Niwat Thepvilojanapong was born in 1977. He received his B.E. degree (Honors) in Electrical Engineering from Chulalongkorn University (Thailand) in 1999. He received his M.E. and Ph.D. degrees in Information and Communication Engineering from the University of Tokyo in 2003 and 2006, respectively. He joined KDDI R&D Laboratories Inc. as a research engineer in 2006. He is currently a project associate professor in the School of Science and Technology for Future Life at Tokyo Denki University from 2008. His research interests include sensor networks and ubiquitous computing. He is a member of the ACM and IEICE.

Masayuki Iwai received Ph.D. in Media and Governance from Keio University in 2002. He is currently a Project Associate Professor of Tokyo Denki University. His research interests include distributed middleware, visual programming, wireless sensor applications, and RFID systems. He is a member of IPSJ.

Oru Mihirogi received B.E. degree in Information Systems and Multimedia Design from Tokyo Denki University in 2007. He is currently a master student at Tokyo Denki University. His research interests include sensor networks and ubiquitous computing. He is a member of IPSJ.

Kazunori Umeda received the B.Eng., M.Eng., and Ph.D. degrees in precision machinery engineering from the University of Tokyo, Japan, in 1989, 1991 and 1994 respectively. He became a Lecturer of Precision Mechanics at Chuo University, Japan in 1994, and is currently a Professor since 2006. He was a visiting worker at National Research Council of Canada from 2003 to 2004. His research interests include range image processing, robot vision, and human-machine interface using vision. He is a member of RSJ, JSPE, JSME, IEICE and IEEE.

Yoshito Tobe received Ph.D. in Media and Governance from Keio University in 2000. He is currently a Professor at Tokyo Denki University. His research includes multimedia communications and sensor networks. He is a member of IEEE Communications Society, ACM, and IPSJ.