歩行者の注目度抽出手法の設計

三尋木 織[†]，焼山 康礼[†]，岩井 将行[††‡]，Niwat Thepvilojanapong[††‡]，
戸辺 義人[††‡]，柴崎 亮介[††]

[†] 東京電機大学 [††] 東京大学 [‡] 独立行政法人科学技術振興機構 CREST

オンラインサービスやアプリケーションの普及により，WWW を利用する人が増加してきている．しかし，商品や展示物のような実世界の物体の注目度を計測する方法が必要である．そこで，実空間における人の注目度を計測するために，我々は a Sensor of Physicalworld Attention using Laser scanning(SPAL) を提案した．レーザスキャナはカメラとは異なり，個人の認識や記録を行なわないため，プライバシーの問題がない．SPAL は注目度を計算するために，停止時間、向き、目標物までの距離などのいくつかの要素を必要とする．注目度を計算するために必要な情報を得るために，我々はレーザスキャナから得た情報を処理する 3 つのモジュールを開発した．我々は，人の注目度を計算するために，2 つの注目指標と 2 つの計測モデルを定義した．システムの有効性を検証するために，我々は SPAL のプロトタイプを実装し，実空間において実験を行った．その結果，提案システムが人の注目度を決定するための有効な手法であることを示した．

Design of Method for Measurement of Pedestrian Attention

Oru MIHIROGI[†], Yasunori YAKIYAMA[†], Masayuki IWAI[††‡], Niwat THEPVILOJANAPONG[††‡], Yoshito TOBE[††‡], Ryosuke SHIBASAKI[††]

[†] Tokyo Denki University [††] University of Tokyo
[‡] CREST, Japan Science and Technology Agency

Human activities in the World Wide Web are increasing rapidly due to the advent of many online services and applications, but 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 Physicalworld 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's behavior. 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 [8, 3]. 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.

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.

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. 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.

2 Related Work

In computer science, *attention* is often referred to as *visual attention* in image processing when recognizing objects. Ma et al. [4] 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. In addition, a human-robot interaction system also requires attention mechanisms to comprehend a situation [1]. 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 [7, 2]. 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 [5]. 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.

3 Human Detection System

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 watch 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. The algorithm is able to extract a human from a group of overlapped human objects. We assume a laser scanner scans in the counter-clockwise direction when briefly describing the algorithm as follows. First, the system generates background data by scanning a measurement area while no one is there. The background data is the longest range of the measurement area. Then margins between the background and the currently scanned data are computed for each scanning step. Beginning edges of a human object are determined according to two cases: (i) the margin is larger than 5cm, (ii) the current data changes more than 20cm in one scanning step. End edges are also determined according to two cases: (i) the margin is smaller than 5cm, (ii) the current data changes more than 20cm. If the distance between the beginning and end edges is longer than 30cm and shorter than 80cm, the recognized ob-

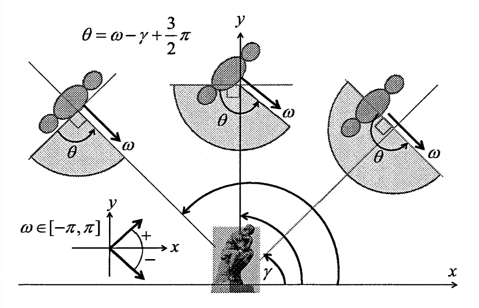


Fig. 1 A angle of human direction (θ).

ject is determined as a human object.

3.2 Detection of Stationary People

We assume moving people are not interested in target objects. Therefore we develop a STOPDETECTION algorithm 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. The difference between two positions ($dist(BeginPosition, EndPosition)$) is the distance that a body covered. If the covered distance is less than or equal to a threshold (30 cm), the person is recognized to be a stationary one.

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.1) to indicate the direction of a human and then calculate people's attention. The angle θ is calculated by the DIRECTIONOFHUMAN algorithm (Fig.2) 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.3). 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. 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).

Algorithm 1 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. 2 DIRECTIONOFHUMAN algorithm.

The 2D vector of the direction of a person is calculated by the DIRECTIONVECTOR algorithm (Fig.3) 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

Algorithm 2 DIRECTIONVECTOR algorithm.

```
procedure DirectionVector(Person, LastAngle, γ)
    // MidPoint: the middle point between BeginEdge and EndEdge
    // V_obj: a normal vector directed to a target object
    // V_opp: a normal vector directed to the opposite direction of a target object
    BeginEdge ← GetBeginEdge(Person)
    EndEdge ← GetEndEdge(Person)
    MidPoint ← FindMidPoint(BeginEdge, EndEdge)
    {V_obj, V_opp} ← FindVector(BeginEdge, EndEdge)
    if (γ − (3/2)π) < LastAngle < (γ − π/2) then
        Direction ← V_obj
    else
        Direction ← V_opp
    end if
    return Direction
end procedure
```

Fig. 3 DIRECTIONVECTOR algorithm.

distance and change in the direction when a person stops.

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.

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 1 expresses the calculation of an instantaneous value of attention (v_{OA}) of the target object o at time t.

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

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.

$$s(o,h,t) = f(\theta(o,h,t)). \qquad (2)$$

The angle θ is measured in the counterclockwise direction from the plane perpendicular to the target object (Fig.1). The function for calculating the direction-based weight is expressed in Equation 3.

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

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 Equation 4.

$$V_{OA}(o, t_a, t_b) = \int_{t_a}^{t_b} v_{OA}(o,t)dt. \qquad (4)$$

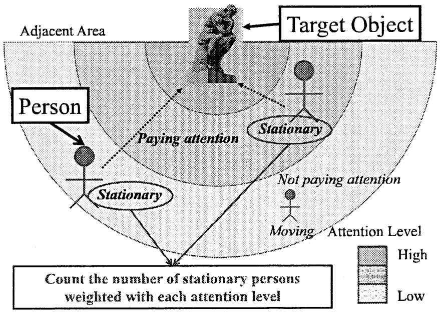


Fig. 4 A distance-weighted attention model.

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 5 and 6 express instantaneous and cumulative degree of real-world attention, respectively.

$$d_{OA}(o,t) = \frac{v_{OA}(o,t)}{N_h(o,t)}. \quad (5)$$

$$D_{OA}(o,t_a,t_b) = \frac{V_{OA}(o,t_a,t_b)}{\int_{t_a}^{t_b} N_h(o,t)dt}. \quad (6)$$

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.4). 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 (Equation 7).

$$v_{DA}(o,t) = \sum_{h=1}^{N_h(o,t)} \{s(o,h,t)\cdot w(o,h,t)\}. \quad (7)$$

The weighting distance is defined in Equation 8.

$$w(o,h,t) = \begin{cases} 1, & dist(o,h,t) \le d_1 \\ \dfrac{d_0 - dist(o,h,t)}{d_0 - d_1}, & d_1 < dist(o,h,t) < d_0 \\ 0, & d_0 \le dist(o,h,t). \end{cases} \quad (8)$$

$dist(o,h,t)$ is the distance between a person h and the target object o at time t. d_0 and d_1 are some parameters to control the effect of distance, where $d_0 \ge d_1$.

An accumulative value of real-world attention can be calculated in the same way as the OA model (Equation 4) 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 (Equations 5 and 6).

For both OA and DA models, if desired, the average value of real-world attention (V) can be calculated by using Equations 9.

$$\overline{V_{\{OA,DA\}}(o,t_a,t_b)} = \frac{V_{\{OA,DA\}}(o,t_a,t_b)}{t_b - t_a}. \quad (9)$$

Table 1 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

5 Implementation and Experiments

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

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[1] in Germany . Table 1 shows the specification of the LMS-200 laser scanner. The laser scanner scans in the counter-clockwise direction. In the SPAL system, we place the laser scanner 140 cm above the ground level. Conse-

[1] Laser Measurement Systems http://www.sick.com/.

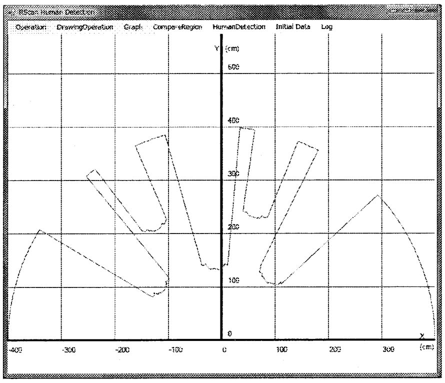


Fig. 5 GUI of application software. The trajectory indicates the front-side circumference of detected objects.

quently, the scanning plane is approximately at the level of the chest of an adult.

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.5) 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.

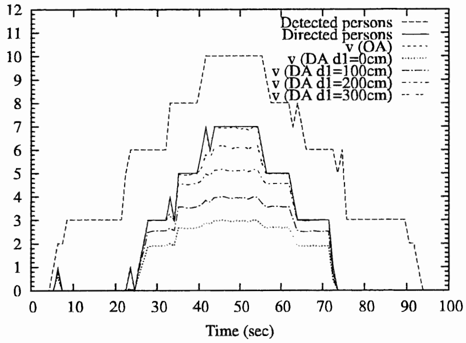
5.2 Experiment Setup

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.

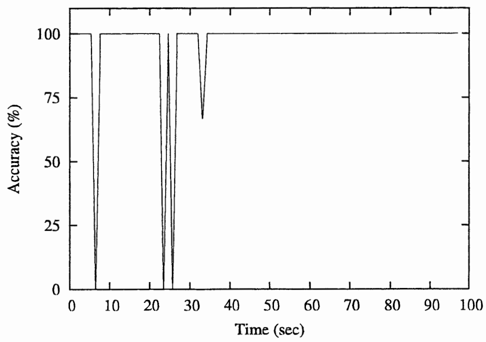
To describe the experiments and results, we define *visitors* and *noises* as follows. The visitors are persons who enter the measurement area and pay attention to the target object, while the noises mean persons who enter the measurement area but they are not interested in the object, i.e., they may 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 the algorithm in Fig.2.

5.3 Experimental Results

Although the experiments were conducted for a long period of time, we show one part of experiments when closely investigating the results. Fig.6 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 at 25 s, 35s and 45s, respectively. Then two, two, and three visitors left at 54, 60, and 70 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. During 5–22 s, three noises were in the measurement area 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 comparable to $Attn_m$ because v_{OA} is weighted by $s(o, h, t)$ (Equation 1) 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 (Equation 7). Therefore, v_{DA} is lower than v_{OA}. As expected, the higher the value of d_1 is, the higher the value of v_{DA}



(a) Instantaneous values of attention (v).



(b) Accuracy of all time intervals.

Fig. 6 Visitors vary from 0–7 persons along the time. Three noises stay in the area all the times.

Table 2 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)
97.09	473	213	210.98	106.90	142.52	178.95	197.74

becomes (Equation 8). 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 Equation 5 decreases while the denominator is still the same. As a result, d always drops when the number of visitors in the area changes. 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 degrees of attention 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} as the results shown in Fig.6(a).

Tables 2 shows cumulative attentions (D) of the above experiments. The attentions are accumulated over the interval shown in Fig.6 (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}) are approaching the number of directed persons, and (ii) the larger d_1 is, the larger the attention becomes.

The accuracy of 6(a) is shown in 6(b). The accuracy of $Attn_m$ is 100% except four steep drops. because the system requires time to detect directed persons. The required time depends on the cycle of measurement (C) used by the StopDetection algorithm(Sect. 3.2). The accuracy of the system is then calculated by Equation 10.

$$\%accuracy = \left(1 - \frac{|Attn_a - Attn_m|}{Attn_a}\right) \times 100 \tag{10}$$

5.4 Discussions

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 two models has been calculated to study the performance of the system. The experimental results show that people's attention correlates 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 events, exhibition halls, museums, etc.

参考文献

1) Imai, M., Ono, T. and Ishiguro, H.: Physical relation and expression: joint attention for human-robot interaction, *IEEE Transactions on Industrial Electronics*, Vol. 50, No. 4, pp. 636–643 (2003).

2) Ishihara, N., Zhao, H. and Shibasaki, R.: Tracking passenger movement with ground-based laser scanner, *Proceedings of the Japan Society of Photogrammetry and Remote Sensing (JSPRS) Annual Conference*, pp. 305–308 (2002).

3) Lee, J. H., Kim, Y.-S., Kim, B. K., Ohba, K., Kawata, H., Ohya, A. and i, S. Y.: Security Door System Using Human Tracking Method with Laser Range Finders, *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation*, pp. 2060–2065 (2007).

4) 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 Transactions on Multimedia*, Vol. 7, No. 5, pp. 907–919 (2005).

5) Nakamura, K., Zhao, H., Shibasaki, R., Sakamoto, K. and Suzukawa, N.: Visualizing passenger flow in railway station using laser scanners, *Proceedings of the 9th International Conference on Computers in Urban Management and Urban Planning* (2005).

6) 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).

7) Zhao, H. and Shibasaki, R.: Pedestrian Tracking using Multiple Laser Range Scanners, *Proceedings of the 8th International Conference on Computers in Urban Planning and Urban Management* (2003).

8) 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, *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, pp. 287–294 (2006).