



Ulm University  
Faculty of Engineering Sciences, Computer Science and Psychology  
Institute of Communications Engineering

## **Study on Human-in-the-loop Sensing in Urban Environment Analysis**

A doctoral thesis jointly supervised with Nara Institute of Science and Technology (Japan)  
and submitted in fulfilment of the requirements for the academic degree of

**Doctor of Natural Sciences**

**Dr. rer. nat.**

by

Yuki Matsuda

born in Kakogawa (Japan)

Acting Dean: Prof. Dr.-Ing. Maurits Ortmanns

Supervisors: Prof. Dr. Keiichi Yasumoto  
Prof. Dr. Dr.-Ing. Wolfgang Minker

Examination date: February 14, 2019

**Doctoral Dissertation**

**Study on Human-in-the-loop Sensing  
in Urban Environment Analysis**

Yuki Matsuda

March 15, 2019

Graduate School of Information Science  
Nara Institute of Science and Technology

A Doctoral Dissertation  
submitted to Graduate School of Information Science,  
Nara Institute of Science and Technology  
in partial fulfillment of the requirements for the degree of  
Doctor of ENGINEERING

Yuki Matsuda

Thesis Committee:

Professor Keiichi Yasumoto	(Supervisor)
Professor Satoshi Nakamura	(Co-supervisor)
Associate Professor Yutaka Arakawa	(Co-supervisor)
Assistant Professor Hirohiko Suwa	(Co-supervisor)
Professor Wolfgang Minker	(Ulm University)

# Study on Human-in-the-loop Sensing in Urban Environment Analysis \*

Yuki Matsuda

## Abstract

Due to the widespread of mobile devices such as smartphones, we can get the appropriate information anytime, anywhere according to the surrounding environments and the human conditions, called “contexts.” In order to provide context-aware systems in urban spaces, it is indispensable to collect information in a complicated and vast spaces comprehensively, and to recognize the context based on the analysis of collected information. Hence, we define the human-in-the-loop sensing framework which consists of two sensing approaches: the direct urban sensing using sensors embedded on mobile devices, and the indirect urban sensing by utilizing humans as a sensor, as the scope of our study. In this dissertation, we focus on two challenges to realize the human-in-the-loop sensing framework: 1) how to extract various context, and 2) how to operate the proposed framework sustainably. Regarding the first challenge, we have tackled with two different cases. One is a safety assessment system for sidewalks at night based on sensing illuminance of streetlamps. We devise a method to correct data from mobile sensors, which have large differences in characteristics and accuracy, utilizing collective intelligence. Through the experiment, we confirmed the proposed method reduced the estimation error for 90 % samples of streetlamps. Another is a psychological context recognition system based on observing the unconscious behavior of tourists. We devise the tourist emotion and satisfaction recognition model by combining multiple modalities collected during sightseeing. With the real-world experiment, we achieved up to 0.50 of average recall score

---

\*Doctoral Dissertation, Graduate School of Information Science, Nara Institute of Science and Technology, March 15, 2019.

with three-class emotion recognition task. As for the second challenge, we build a mobile participatory sensing platform, which incorporates citizen communities into the ecosystem of human-in-the-loop sensing framework. Especially, we focus on “civic tech” which is one of the ways to realize civic cooperation by using ICT, and have placed civic tech communities as the potential user group of our platform. Through 17 case studies with citizens, we have confirmed the usability and availability of the proposed platform.

**Keywords:**

context awareness, smart city, participatory sensing, ubiquitous computing, mobile computing, human-in-the-loop.

# Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background and Scope of Study . . . . .	1
1.2 Human-in-the-loop Sensing . . . . .	4
1.3 Challenge Statements . . . . .	6
1.4 Organization of Dissertation . . . . .	8
<b>2 NightRoadScanner: Urban Environment Analysis Using Mobile Sensors</b>	<b>9</b>
2.1 Introduction - Background and Motivation . . . . .	9
2.2 Preliminary . . . . .	10
2.2.1 Literature Review . . . . .	10
2.2.2 Previous Work and Achievements . . . . .	12
2.2.3 Remaining Challenges of Study . . . . .	15
2.3 Methodology of Streetlamp Illuminance Estimation . . . . .	17
2.3.1 Concept of Illuminance Data Correction Method . . . . .	17
2.3.2 Correction of Obliquely Incident Light Characteristic . . . . .	19
2.3.3 Correction of Illuminance Following Characteristic . . . . .	22
2.3.4 Streetlamp Illuminance Estimating Method . . . . .	24
2.3.5 Measures Against Propagation Error of Correction . . . . .	26
2.4 Experiments and Evaluation . . . . .	28
2.4.1 Overview of Experiments and Evaluation Metrics . . . . .	28
2.4.2 Evaluation Result and Discussion . . . . .	31
2.5 Conclusion . . . . .	36

<b>3</b>	<b>EmoTour:</b>	
	<b>Urban Environment Analysis Using Human as a Sensor</b>	<b>37</b>
3.1	Introduction - Background and Motivation . . . . .	37
3.2	Literature Review and Challenges . . . . .	38
3.2.1	Estimation of Emotional Status . . . . .	39
3.2.2	Estimation of Satisfaction Level . . . . .	40
3.2.3	Objective and Challenges . . . . .	41
3.3	Proposed Approach and Workflow . . . . .	41
3.3.1	Modalities . . . . .	43
3.3.2	Labels . . . . .	44
3.4	Methodology of Tourist Emotion and Satisfaction Estimation . . .	46
3.4.1	Preprocessing and Feature Extraction . . . . .	47
3.4.2	Modality Fusion . . . . .	51
3.5	Experiments and Evaluation . . . . .	52
3.5.1	Overview of Real-World Experiments . . . . .	52
3.5.2	Results . . . . .	56
3.6	Discussion and Limitations . . . . .	58
3.6.1	Feasibility of Proposed System . . . . .	58
3.6.2	Imbalance of Labels . . . . .	58
3.6.3	Limitation of Data Sources . . . . .	60
3.6.4	Future Perspectives . . . . .	61
3.7	Conclusion . . . . .	61
<b>4</b>	<b>ParmoSense: A Pariticipatory Mobile Sensing Platform</b>	<b>62</b>
4.1	Introduction - Background and Motivation . . . . .	62
4.2	Literature Review and Challenges . . . . .	63
4.2.1	Functions Essential in Participatory Sensing Platform . . .	63
4.2.2	Required Skills for Operation and Use . . . . .	68
4.3	Scenario-based Participatory Sensing Platform . . . . .	70
4.3.1	ParmoSense Basic Principles . . . . .	71
4.3.2	ParmoSense System Overview . . . . .	72
4.3.3	ParmoSense System Architecture . . . . .	74
4.3.4	Functions Support . . . . .	80

4.4	Evaluation . . . . .	83
4.4.1	Workload for System Operation . . . . .	83
4.4.2	Variety of Functions . . . . .	84
4.5	Case Studies . . . . .	86
4.5.1	Overview of Case Studies . . . . .	88
4.5.2	Survey and Discussion . . . . .	92
4.6	Conclusion . . . . .	99
<b>5</b>	<b>Conclusion</b>	<b>101</b>
5.1	Summary . . . . .	101
5.2	Discussion and Future Perspective . . . . .	104
5.2.1	Privacy Considerations . . . . .	104
5.2.2	Knowledge Reusability . . . . .	104
5.2.3	Method Generalizability . . . . .	105
5.2.4	Considerations of Personality and Preference . . . . .	105
	<b>Acknowledgements</b>	<b>107</b>
	<b>References</b>	<b>110</b>
	<b>Appendix</b>	<b>124</b>
A.	Derivation of Illuminance Model at Walking . . . . .	124
B.	Correction of Smartphone Holding Angle . . . . .	126
C.	Sample Scenario of ParmoSense . . . . .	128
	<b>Glossary</b>	<b>129</b>
	<b>Publication List</b>	<b>141</b>

# List of Figures

1.1	The human involvement model in urban sensing system. . . . .	2
1.2	Human-in-the-loop sensing, an urban sensing framework for collecting both objective and subjective contexts of urban environments, consisting of the two involvement levels: <i>human as a device carrier</i> and <i>human as a sensor</i> . . . . .	5
1.3	Challenge Statements and Organization of Dissertation. . . . .	7
2.1	CrimeReports.com [13]. . . . .	10
2.2	An example of mapping brightness of road at night [16]. . . . .	11
2.3	Illuminance measurement points based on the illuminance measurements method which Japanese industrial standards committee (JISC) defined [18]. . . . .	12
2.4	The workflow of inferring safety level of sidewalks at night utilizing smartphone embedded light sensors. . . . .	13
2.5	An illuminance transition model while walking. The parameter $\theta_s$ represents the angle of smartphone in human's hand. . . . .	14
2.6	The incidence angle characteristics of the illuminance of Nexus 7, Galaxy Nexus. . . . .	16
2.7	A correlation chart of illuminance (illuminometer vs Nexus 7, Galaxy Nexus). . . . .	16
2.8	The concept of the relative correction method of measured street-lamp's illuminance at a unit area. . . . .	18
2.9	The concept of the relative correction method of measured street-lamp's illuminance at all areas. . . . .	18
2.10	Correction parameters of an incident angle characteristics of illuminance. . . . .	20

2.11	An adverse effect of a user’s shadow. . . . .	23
2.12	An adverse effect of an accidental light incident. . . . .	23
2.13	The inference of illuminances on the centerline of the road and the opposite line of the road. . . . .	24
2.14	The inference of illuminances on the centerline of the road. . . . .	25
2.15	The history tree of the correction. . . . .	27
2.16	The reduction of the accumulative error of propagations by labeling of the number of times applying the correction method. . . . .	27
2.17	The evaluation field of the correction method. . . . .	29
2.18	The result of the fitting method of illuminance transition. . . . .	32
2.19	The comparison of illuminance transitions on streetlamp No. 2. . . . .	33
2.20	The comparison of illuminance transitions on streetlamp No. 12. . . . .	34
2.21	The result of illuminance estimation method with corrected data. . . . .	35
3.1	Workflow to estimate tourist emotional status and satisfaction level. . . . .	42
3.2	Devices for data collection during sightseeing. . . . .	43
3.3	Smartphone application for collecting labels from tourists. . . . .	45
3.4	A two-dimensional emotion status model. This figure is derived from [58, 59]. . . . .	45
3.5	Scheme of tourist emotion and satisfaction estimation with modal- ity fusion. . . . .	46
3.6	Spherical coordinates of pupil. . . . .	47
3.7	Threshold of eye movement ( <i>theta</i> and <i>phi</i> ). . . . .	47
3.8	Video features ( <i>Action Units</i> ) extraction using <i>Openface</i> [63, 64]. . . . .	50
3.9	Experimental fields. . . . .	53
3.10	The distribution of labels (emotion). . . . .	54
3.11	The distribution of labels (satisfaction). . . . .	54
3.12	The relationship between emotional status and satisfaction level. . . . .	55
4.1	Design concept . . . . .	73
4.2	System architecture of ParmoSense . . . . .	75
4.3	Internal system (Figure 4.2 ⑤) . . . . .	75
4.4	Scenario tools interface of ParmoSense Dashboard (Figure 4.2 ①) . . . . .	76
4.5	Data tools interface of ParmoSense Dashboard (Figure 4.2 ②) . . . . .	77

4.6	ParmoSense Client interface . . . . .	78
4.7	Examples of motivating functions . . . . .	81
4.8	Comparison graph . . . . .	86
4.9	Result of human-behavior investigation . . . . .	92
A.1	A positional relationship between a streetlamp and a smartphone embedded light sensor. . . . .	125
A.2	Obliquely incident light (OIL) characteristic of Galaxy Nexus. . .	126
B.1	The correction of the holding angle of a smartphone. . . . .	127

# List of Tables

2.1	The illuminance standard for streetlamps [19] Japan security systems association (JSSA) defined. . . . .	12
3.1	Performance of uni- and multimodal tourist emotion and satisfaction estimation. . . . .	56
3.2	Performance of tourist emotion and satisfaction estimation (by nationality of participants). . . . .	59
4.1	Overview of supported functions . . . . .	64
4.2	Overview of platform skill requirements. . . . .	69
4.3	Preparation Time for Each Task . . . . .	85
4.4	Breakdown of Preparation Time . . . . .	85
4.5	Overview of case studies. . . . .	87
4.6	Questionnaire items for participants and non-participants (S2, S4–S6, S9, S11–S14) . . . . .	94
4.7	Questionnaire items for organizers (S2, S4–S6, S9) . . . . .	97

# 1 Introduction

## 1.1 Background and Scope of Study

Along with the great development of information and communication technology (ICT) in the 21st century, there are more than five billion users of mobile devices including smartphones and wearable devices, i.e., such devices are becoming indispensable to our lives [1]. Such technologies which have been evolved to a high level and spread in daily life, called ubiquitous computing [2], have a potential to provide useful information to people anytime and anywhere based on sensing and analyzing environments. Indeed, thanks to technological improvements in smartphones, Internet of Things (IoT), mobile networks, and others, we can get the appropriate information depending on the situation of the surrounding environment (objective context), even when people are outdoors such as urban environments. For example, Nishimura et al. have proposed a method to estimate the traveling situation of a train from sensors embedded on smartphones owned by general people to avoid congestion on a train and minimize the waiting time of passengers [3]. Morishita et al. have proposed a system that recognizes and recommends routes with beautiful cherry blossoms based on videos recorded by a car-mounted video camera such as a drive recorder [4]. In recent years, in addition to the state of the urban environment, the understanding of the mental state (subjective context) such as emotion and satisfaction which human felt attracts attention. Hence, the more context-aware system which can provide information based on understanding the subjective context of people in the urban space might be demanded in the future. The aim of this study is to collect and recognize both objective and subjective contexts in the urban environment.

In order to provide context-aware systems in urban spaces, it is essential to comprehensively collect information in a complicated and vast urban spaces, and

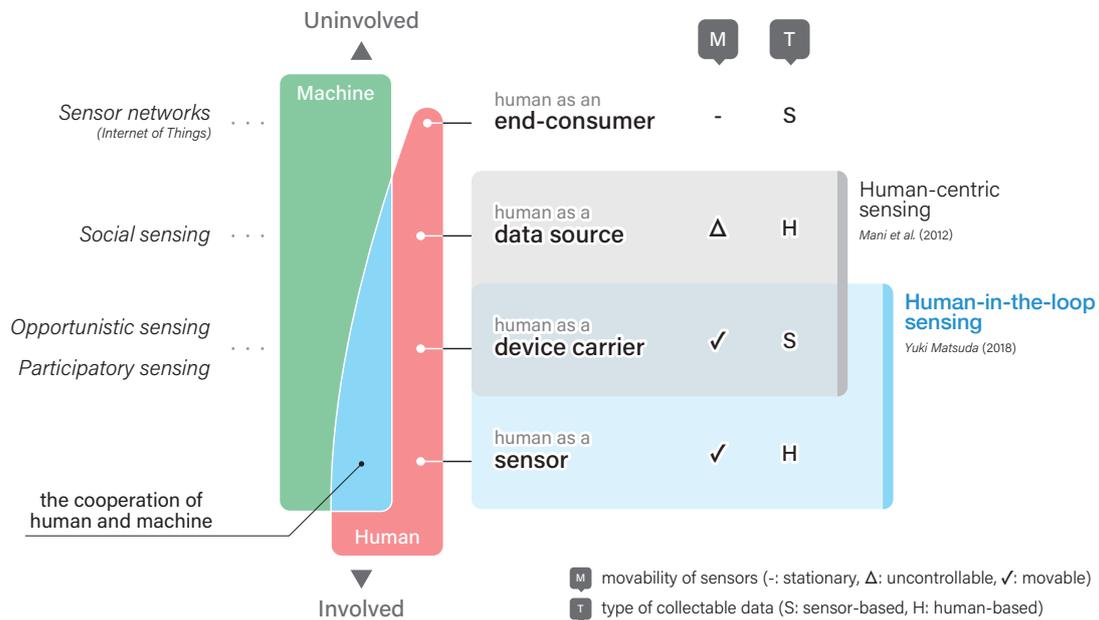


Figure 1.1. The human involvement model in urban sensing system.

to recognize the contexts based on the analysis of collected information. These issues have become one of the important research topics in the smart city domain, and many studies have been done. In general, to realize the data collection and context recognition in urban spaces, the following problems should be solved:

- **Lack of recognizable context variety**

In the real-world, especially the urban environment, there are countless contexts and combinations to be recognized.

- **Limit of spatiotemporal data coverage**

Due to the vast space in the urban environment, it is hard to collect enough amount of data in every place and time.

To solve these problems, various approaches have been proposed and put into practical use. We organized each approach in the viewpoint of human involvement in urban sensing systems as shown in Figure 1.1 (named human involvement model). The vertical axes indicate the extent of human involvement, which is increased toward lower layers. Also, regarding each involvement level layer, the

following two attributes are indicated: the *movability of sensor* (i.e., controllability of data collection point), and the *type of collectible data*. In the research field of engineering in recent years, the emphasis is placed on unmanned and automated by IoT and a sensor network utilizing it, and tasks that machines can substitute tend to be excluded from people’s hands (***Human as an end-consumer***). With this approach, the efficiency of data collection is dramatically improved, and continuous data collection became possible for places where sensors were installed. However, as pointed out by Mani et al., the sensor networks have the following problems [5]: the limited coverage of spatiotemporal data (data can be acquired only in places where sensors are present); the limit of making complicated reasoning, and difficulty to be applied to crowded places (the recognizable context is limited to the range observable by the sensor).

From this background, Mani et al. have proposed a sensing method, named ***human-centric sensing***, based on the cooperation of human and machines by involving humans to the sensing process [5]. The components of *human-centric sensing* are outlined below<sup>\*1</sup>. The first is social sensing [6], a method for extracting the various contexts from posts in social networking services (SNSs). By analyzing the SNSs data such as natural languages, photographs, and movies posted by general people, this enables us to collect both objective and subjective contexts that can not be measured by sensors but can be recognized by the human (i.e., this approach regards ***human as a data source***<sup>\*2</sup>). The second is participatory sensing [7], a method based on requesting people, who exist in the environment, to contribute to urban sensing. By using sensors embedded in mobile devices owned by general people, it is possible to collect information equivalent to sensor networks without needing infrastructure. In addition, it is also possible to collect spatially or temporally insufficient data by daringly asking people (i.e., this approach regards ***human as a device carrier***<sup>\*3</sup>). In summary, involving *human* into the system has a potential to solve the problems which the conventional automated sensing system is facing.

---

<sup>\*1</sup> While *human as targets of sensing* which focuses lifelog is also included in the definition of *human-centric sensing*, it is deviated from urban environment sensing domain. Hence it is not plotted in Figure 1.1.

<sup>\*2</sup> In the definition of Mani et al., it is defined as *human as data sources*.

<sup>\*3</sup> In the definition of Mani et al., it is defined as *human as sensor operators*.

As mentioned above, in order to realize the next generation context-aware system, we must consider not only the objective context of the urban environment but also the subjective context such as how the people feel in the urban space. In other words, a combination of direct measurement using sensors and indirect measurement using the human is essential (the combination of *human as a data source* and *human as a device carrier* is employed in *human-centric sensing*). However, *human as a data source* is a mechanism based on diverting a huge amount of SNS data posted by people who are not conscious of the contribution to the urban sensing system. It means the involvement level of people is low, and the system cannot control people’s behavior directly. It suggests there is a potential issue that spatiotemporal bias occurs in collected data, and it is difficult to improve this.

Thus, we newly defined ***human as a sensor*** as a deeper level of human involvement. *Human as a sensor* is a sensing approach that utilizes a sensory system of human (e.g., five senses) and psychological reactions (e.g., emotional status) as a new urban environmental sensor into the system. Then, we have defined a new urban environmental sensing framework, named ***human-in-the-loop sensing***, consisting of the two involvement levels: *human as a device carrier* and *human as a sensor*. We positioned this framework as a scope of our research. In the next section, *human-in-the-loop sensing* will be described in detail.

## 1.2 Human-in-the-loop Sensing

The concept of *human-in-the-loop sensing*, a scope of this study, is shown in Figure 1.2 and explained in detail below.

*Human-in-the-loop sensing* is based on the crowdsourcing-based sensing method in order to comprehensively collect and analyze various contexts in the urban environment. This method collects the urban environmental information by using the sensors embedded in the smartphones or wearable devices owned by general people in the urban space. Regarding this method, it is roughly divided into two categories: participatory sensing, the user intentionally performs sensing tasks based on the request; opportunistic sensing, the system implicitly performs sensing tasks after getting user’s agreement.

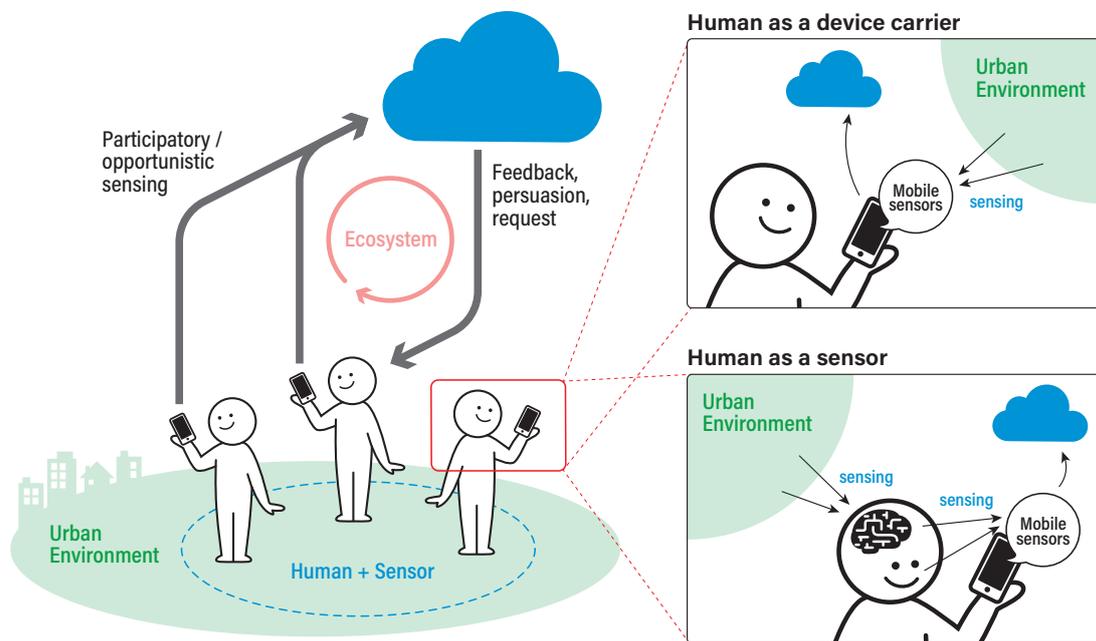


Figure 1.2. Human-in-the-loop sensing, an urban sensing framework for collecting both objective and subjective contexts of urban environments, consisting of the two involvement levels: *human as a device carrier* and *human as a sensor*.

In addition, *human-in-the-loop sensing* employs the sensing method at the following two involvement levels among the human involvement model shown in the Figure 1.1:

***Human as a device carrier*** is a concept based on the sensing method in which human is involved in the system as “mobile sensor carriers,” and mobile sensors directly sense the environment. Many of the conventional participatory sensing and opportunistic sensing correspond to this. The main sensing target is urban environmental information that can be directly detected by the mobile sensor, such as the sound noise [8, 9], the mobile network condition [10], the air pollution [11, 12]. Additionally, the input interfaces, e.g., a camera and a microphone, are regarded as sensors, and media data from them such as photos and videos [4] are also utilized as sensor data.

***Human as a sensor*** is a concept based on the sensing method in which sensory systems of human are involved in the system as “urban environmental sensor,” and indirectly sense the environment by extracting reaction of the human. By delegating the role of sensors in *human as a device carrier* concept to humans, the human is involved at a deeper level in the system. The main sensing target is the phenomenon that the human can perceive, and information about how the human feels in the urban space. In order to acquire reactions of the human, a smartphone or a wearable device is employed.

Since sensing methods based on these two involvement levels can not completely substitute each other, selecting or combining these concepts is needed to recognize various contexts existing in the urban environment.

### 1.3 Challenge Statements

The urban environmental sensing system based on *Human-in-the-loop sensing* gives us the following benefits:

- A sensing network can be easily constructed by utilizing the hundred millions of mobile devices owned by general people and the mobile network infrastructure spread around the world.
- The mobile network and power supply of the mobile device are always managed and maintained in good condition by each user.
- Thanks to human mobility, it is possible to improve the spatiotemporal coverage of sensing data.
- Based on selection and integration of mechanical sensors and human sensors, it is possible to understand various contexts.

However, since *human-in-the-loop sensing* is a framework established by the deeper involvement of the human, it has its inherent problems. We organize them as challenges of this study in Figure 1.3 and describe as follows:

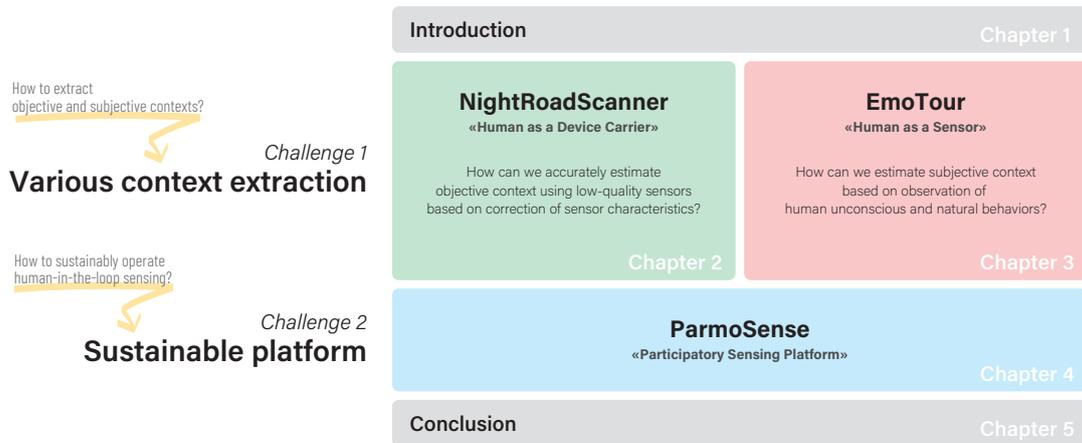


Figure 1.3. Challenge Statements and Organization of Dissertation.

### Challenge 1: How to extract objective and subjective contexts?

*Human-in-the-loop sensing* employs data collection mechanisms using common mobile devices such as smartphones and wearable devices which have various sensors with greatly different characteristics and accuracy, or human as a sensor which is extremely ambiguous existence. Hence, the key point is how to extract useful information from various low-quality data collected by them. In this study, we tackle the following challenges about *human as a device carrier* and *human as a sensor*. Regarding *human as a device carrier*, we pick up the sample case about safety assessment method for the sidewalks at night based on measurement of the illuminance of streetlamps. The light sensor embedded on smartphones has a problem that the sensor characteristics differ not only between different models, but even between the same models. Hence, we consider a method to relatively correct data measured by unknown devices using the data measured by devices whose sensor characteristics is known. Regarding *human as a sensor*, we pick up another sample case of estimation method of the subjective contexts (emotional status and satisfaction level) by observing unconscious behaviors and reactions of tourists during sightseeing. Human subjective context has been collected only by questionnaire or user review so far, however, such information might be demanded in the future context-aware systems. Thus, we

consider a method to estimate them based on the behavior naturally and unconsciously performed by the human such as body and head movement, facial- and vocal-expressions.

**Challenge 2: How to sustainably operate *human-in-the-loop sensing*?**

*Human-in-the-loop sensing* is a mechanism established on “cooperation” in which depends on the voluntarism of general people. In other words, the sustainability of the system is a critical challenge in real-world operation. As an idea to enhance this sustainability, the mutual linkage with the local community where ecosystems are already formed can be considered. For example, the civic cooperation that people work with government, universities, companies, etc. to promote community development spreads globally. Especially in recent years, *civic technology* (*civic tech*) which combines ICT and civic cooperation is gathering attention. We focus on the potential of a high synergy produced by civic tech and *human-in-the-loop sensing*. Hence in this study, we design the *human-in-the-loop sensing* platform for easy construction and operation of urban environment sensing systems in the civic tech community.

## 1.4 Organization of Dissertation

This dissertation is organized as follows. In Chapter 2, as the objective context estimation method based on *human as a device carrier* concept, we explain a method of correcting mobile sensors which have different characteristics and accuracy, and estimating urban environmental information. Then, in Chapter 3, as the subjective context estimation method based on *human as a sensor* concept, we describe a method of estimating psychological state (emotional status and satisfaction levels) based on observation of unconscious behavior of people. In addition, Chapter 4 provides a design and implementation of the mobile sensing platform for operating *human-in-the-loop sensing* as a sustainable framework in the real-world condition. Finally, Chapter 5 concludes this dissertation.

## 2 NightRoadScanner: Urban Environment Analysis Using Mobile Sensors

In this chapter, as the sample case of sensing approach based on the *human as a device carrier* concept, we describe a study on a safety level assessment method of sidewalks at night using participatory illuminance sensing system with common smartphones.

### 2.1 Introduction - Background and Motivation

In recent years, safety and security awareness of street crime is growing worldwide. The information on safety or security, e.g., crime information, the safety level of sidewalks at night and disaster information, should be provided by next generation context-aware systems in addition to common environmental information.

Such security information is collected and visualized on CrimeReports [13] as shown in Figure 2.1. The service plots details and locations of crimes published by law enforcement agencies on the map. However, we have to estimate safe areas manually, because we can not grasp the safety level of a local area by only glancing at this plotted information.

Apart from that, a local Japanese organization has created a safety assessment tool for sidewalks at night based on streetlamps' illuminance measured with an illuminometer. According to the report from the Japan security systems association (JSSA) [14], the illuminance of the streetlamps can contribute to security improvement. It suggests it is possible to judge a safety level easily and visually.

Figure was removed due to copyright reasons.

Figure 2.1. CrimeReports.com [13].

However, a complete survey of street lighting with an illuminometer does not seem practical due to equipment and labour costs.

Here, we have introduced the *human-in-the-loop sensing* framework, especially the *human as a device carrier* concept, to safety level assessment via illuminance of streetlamps in order to address this problem. In the following sections, we explain a participatory illuminance sensing system, named NightRoadScanner [15], for safety level assessment system of sidewalks at night.

## 2.2 Preliminary

### 2.2.1 Literature Review

As a way of collecting the brightness of the night road, there is a method of estimating via the collection and analysis of the information of streetlamp installed by each municipality (such as open data). However, in the real-world, there is a problem that such data is not organized due to the jurisdiction of streetlamp installation is scattered to various organizations, e.g., the country, prefecture, city, town and furthermore individual. In addition, there is also a problem that the



Figure 2.2. An example of mapping brightness of road at night [16].

illuminance of streetlamps should be regularly surveyed, since the aged deterioration of streetlamps might affect safety levels.

Yamada et al. have proposed the method to assess the safety level of sidewalks at night based on measuring illuminance of streetlamps [16,17]. Figure 2.2 shows an example of mapping measured values. As a method to assess the safety level of sidewalks at night, they have employed the illuminance measurements method [18] which Japanese industrial standards committee (JISC) have defined (hereinafter called *JIS method*), and the illuminance standard [19] shown in Table 2.1 which JSSA have defined<sup>\*1</sup> (hereinafter called *JSSA standard*). Figure 2.3 shows an outline of the JIS method. The horizontal (vertical) illuminance is measured by a horizontally (vertically) held illuminometer against to the road surface (road direction) at each measurement points. JIS method requires that each measurement point is placed as mesh-pattern which have 5 m interval, i.e., we need to measure illuminance at approximately 10–50 measurement points for every streetlamp. This approach has a problem in difficulty of keeping data coverage due to labor and equipment costs.

<sup>\*1</sup> This standard is designed according to guidelines for streetlamp installation [20] defined by JISC.

Table 2.1. The illuminance standard for streetlamps [19] Japan security systems association (JSSA) defined.

Class	A (more safe)	B (safe)
Average value of horizontal illuminance <sup>*a</sup>	over 5.0 lx	over 3.0 lx
Minimum value of vertical illuminance <sup>*b</sup>	over 1.0 lx	over 0.5 lx
Illumination effect <sup>*c</sup>	Recognizable the outline of the opponent face	Recognizable the posture and the behavior

<sup>\*a</sup> The average illuminance on the road surface at height of 0 m.

<sup>\*b</sup> The minimum illuminance of the vertical plane against the road direction at height of 1.5 m.

<sup>\*c</sup> The metric showing how much information of a person in 4 m ahead can be obtained.

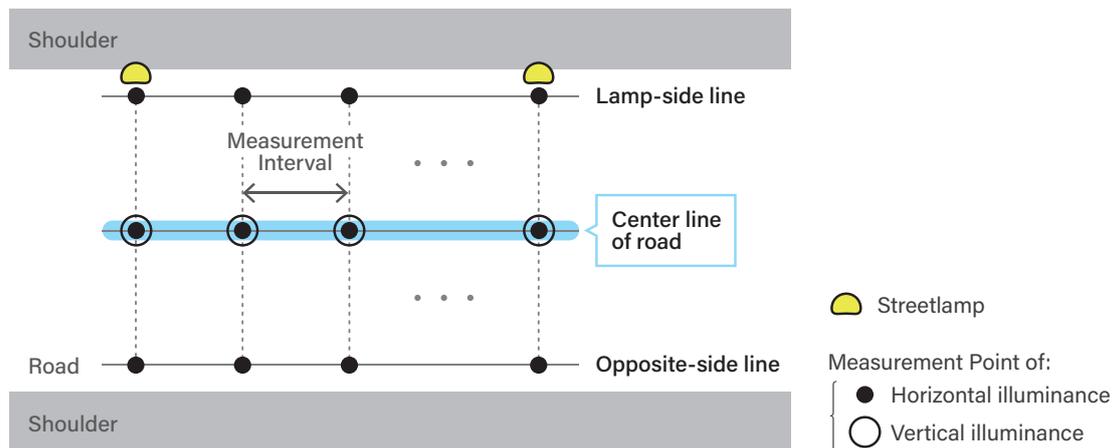


Figure 2.3. Illuminance measurement points based on the illuminance measurements method which Japanese industrial standards committee (JISC) defined [18].

## 2.2.2 Previous Work and Achievements

We have studied the method to realize easily measuring illuminance of streetlamps according to JIS method by estimating the position, type, and illuminance of streetlamps using data measured with smartphone embedded light sensors which general people have, in our previous work [21–24]. The workflow of this approach is shown Figure 2.4 and described as follows:

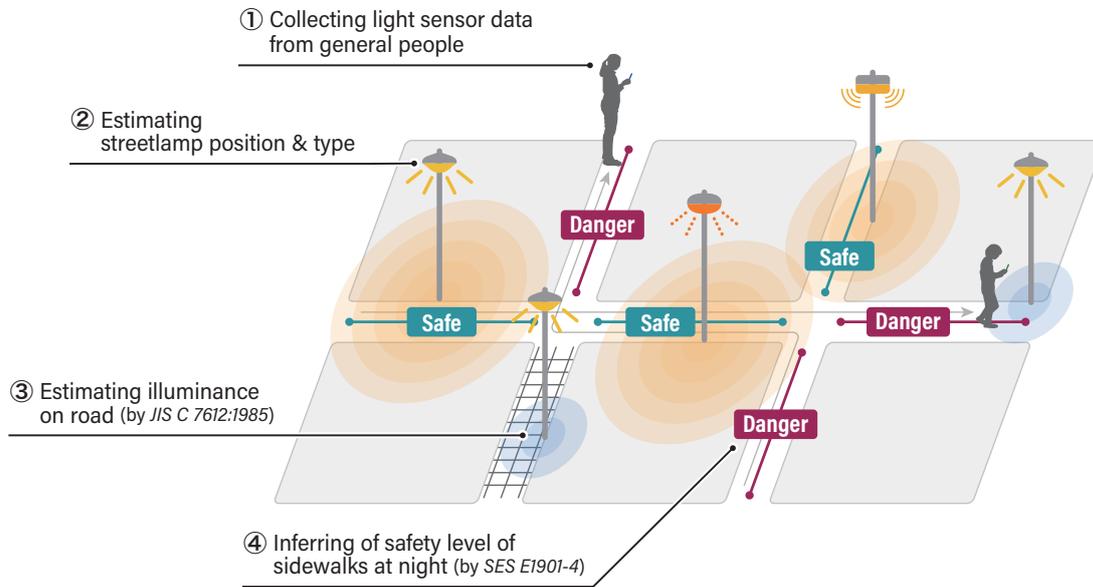


Figure 2.4. The workflow of inferring safety level of sidewalks at night utilizing smartphone embedded light sensors.

### Step 1: Collecting light sensor data from general people [21]

At first, sensor data (e.g., illuminance, acceleration, GPS) are gathered from smartphones of an unspecified number of users. Since these data are measured in various conditions such as different device model, device holding angle and walking speed, illuminance data might include noises. Hence, by estimating the measurement condition using other sensor data, illuminance data is converted in order to it becomes close to data measured using illuminometer.

### Step 2: Estimating streetlamp position and type [22, 23]

Since JIS method requires the measurement of illuminance between streetlamps, the distance between two adjacent streetlamps is an essential parameter to infer the safety level of sidewalks at night. Furthermore, JSSA defines a different illuminance standard for LED lights, thus we need to distinguish a type of streetlamps. However, data sets or databases which have such information is not provided as mentioned above. Hence, we estimate

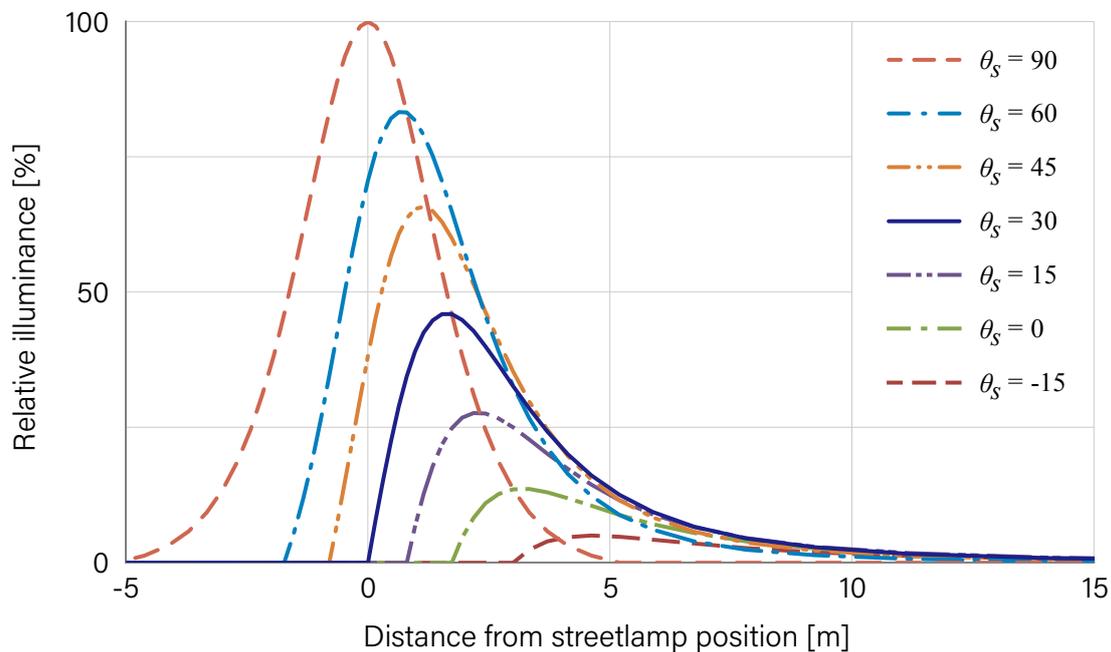


Figure 2.5. An illuminance transition model while walking. The parameter  $\theta_s$  represents the angle of smartphone in human’s hand.

the position and type of streetlamps based on the comparison between the measured sensor data using multiple smartphone embedded sensors and the illuminance model at walking<sup>\*2</sup>.

### Step 3: Estimating illuminance on road [24]

Then, we estimate illuminance transitions on center, lamp-side, opposite-side line from measured illuminance data. We calculate the illuminance values according to JIS method [18] using mesh-pattern illuminance data.

### Step 4: Inferring safety level of sidewalks at night [24]

Finally, the safety level of sidewalks at night is assessed with comparing estimated illuminance value and JSSA standard as shown in Table 2.1.

<sup>\*2</sup> The model of illuminance transition when a user passes just under a streetlamp with smartphone held at angle  $\theta_s$ . The procedure for deriving illuminance model at walking is described in Appendix A..

### 2.2.3 Remaining Challenges of Study

Illuminance is generally known to be sensitive to distance and angle from a light source. In addition, whereas the light receiver of illuminometer has a hemispherical shape, the smartphone embedded light sensor has a planar shape because it is installed inside the device. From this fact, the smartphone embedded light sensor has specific characteristics different from illuminometer. Hence, in the previous work [21–24], we have proposed each method on the assumption that the characteristic between the same device model is the same.

However, we have not verified whether this assumption holds for any other models. To investigate this, we have conducted a preliminary experiment using four smartphones of the same device model. In the experiment, we check the following characteristics: obliquely incident light (OIL) characteristic, the characteristic of illuminance decay caused by angle of incident light; and illuminance following (IF) characteristic, the characteristic of the tracking performance to illuminance intensity changing. We used an illuminometer (model: LX-1108, manufacturer: Satoshoji Corp., measurement range [resolution]: 0–40.00 [0.01] lx, 36.0–400.0 [0.1] lx) and tablet devices (model: Nexus 7 ME571-16G, manufacturer: Samsung, measurement range [resolution]: 0–32000 [0.01] lx)<sup>\*3</sup>.

Figure 2.6 shows the survey result of OIL characteristic. While this characteristic follows Lambert’s cosine law in general, the cosine to the 5th power ( $\cos^5$ , it shows a stronger attenuation) for smartphone embedded light sensor. This attenuation is different depending on each devices, there is up to 15 % difference at near incident angle of 90 degree. Then, Figure 2.7 shows the survey result of IF characteristic. Here, the incident angle has been set as 90 degree. We confirmed each device has proportional characteristic, however the coefficient value is definitely different (there is up to 50 % difference).

From this results, since there is large different in both OIL characteristic and IF characteristic even same device model, we can not apply the same data correction method. *Human-in-the-loop sensing* assumes the measurement using light sensors embedded on various smartphones which an unspecified number of people has. Hence, the correction method which can be applied any kind of sensors is essential.

---

<sup>\*3</sup> All the devices were cleaned and not covered with any screen protection film and used in the same environment.

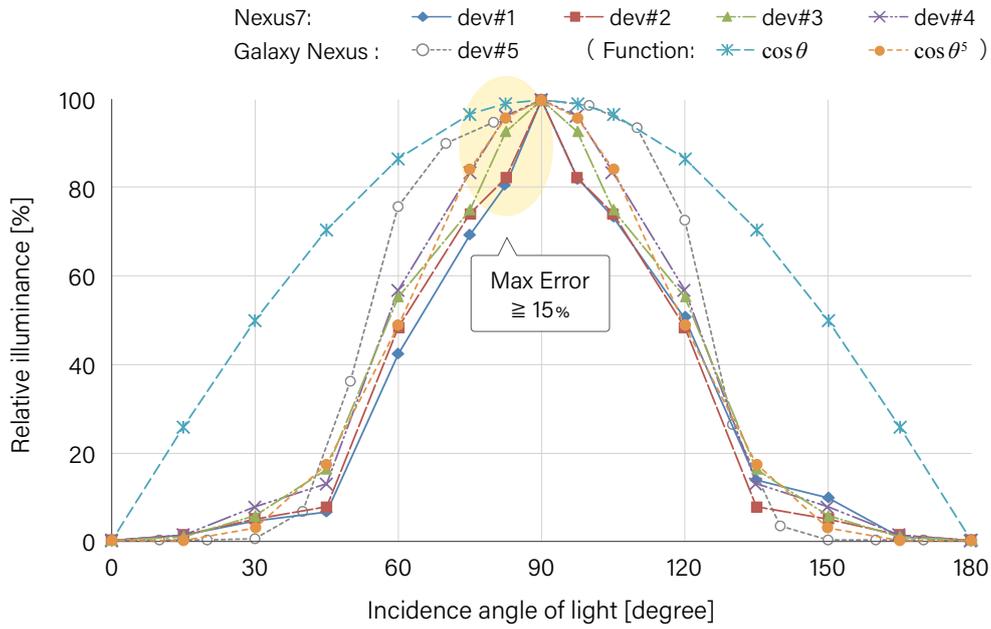


Figure 2.6. The incidence angle characteristics of the illuminance of Nexus 7, Galaxy Nexus.

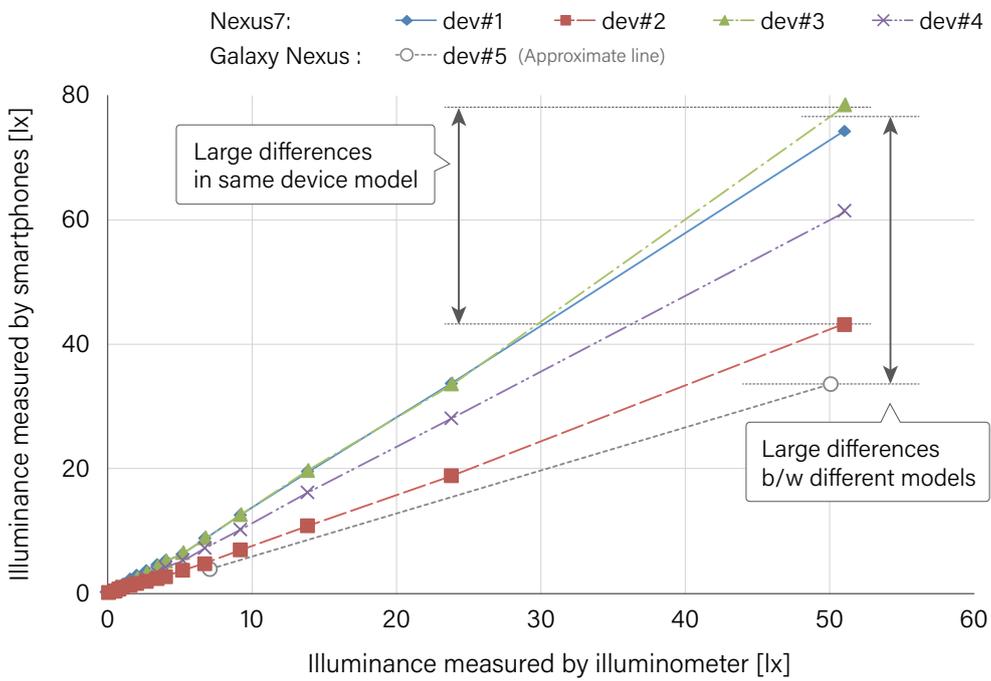


Figure 2.7. A correlation chart of illuminance (illuminometer vs Nexus 7, Galaxy Nexus).

## 2.3 Methodology of Streetlamp Illuminance Estimation

### 2.3.1 Concept of Illuminance Data Correction Method

To realize sensing using devices with different characteristics, we have been used the model-by-model correction approach, in the previous work. However, from the results of the survey in Section 2.2.3, it turned out that the current correction approach was not enough. On the other hand, it is not realistic that we investigate characteristic with all available devices (even the same model), from the perspective of data coverage.

Hence, we propose the method of relative correction of illuminance data. The steps are described as follows:

1. We assume there is the following two kind of users (devices) shown in Figure 2.8: the correction standard, its sensor characteristic is known in advance (User A); the correction target, its sensor characteristic is unknown (User B). We correct data of User B with comparing illuminance data for the same streetlamp (Streetlamp 1) measured by both Users A and B.
2. The parameters to correct both illuminance characteristics (OIL characteristic, IF characteristic) between User A and B are extracted. Section 2.3.2 and 2.3.3 provide detail explanation of each method. The illuminance data for other streetlamps are also corrected using extracted parameters.
3. In other area, corrected device (User B) will become new correction standard, and correct other devices as shown in Figure 2.9. To avoid a propagation errors, we consider the method described in Section 2.3.5.
4. The comprehensive and relative correction network will build by repeating procedure 1 to 3. Section 2.3.5 provides a detail explanation.

As our previous work, the illuminance estimation method based on JIS method utilizing smartphone embedded light sensors whose sensor characteristics are already-known has been proposed [24]. This chapter extends this method using additionally light sensors whose sensor characteristics is unknown.

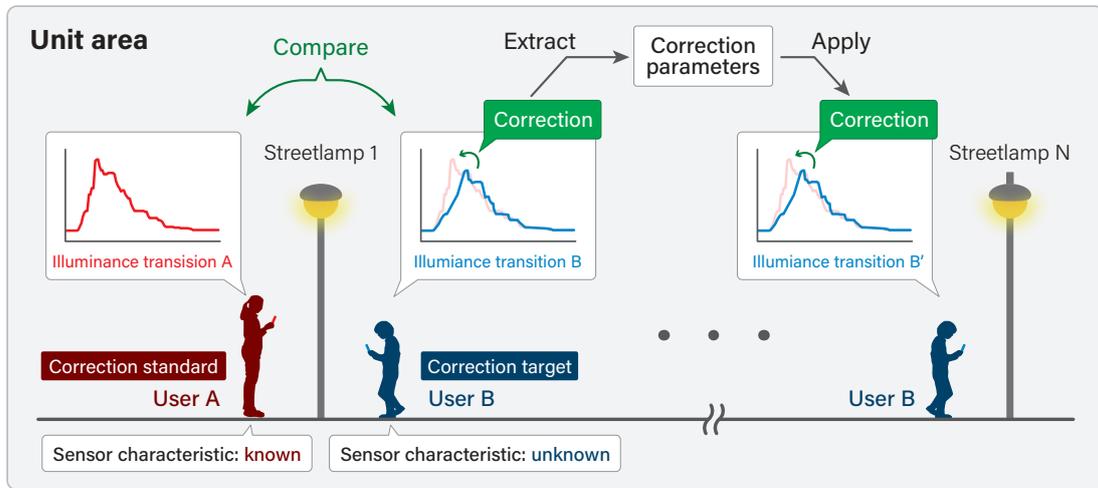


Figure 2.8. The concept of the relative correction method of measured street-lamp's illuminance at a unit area.

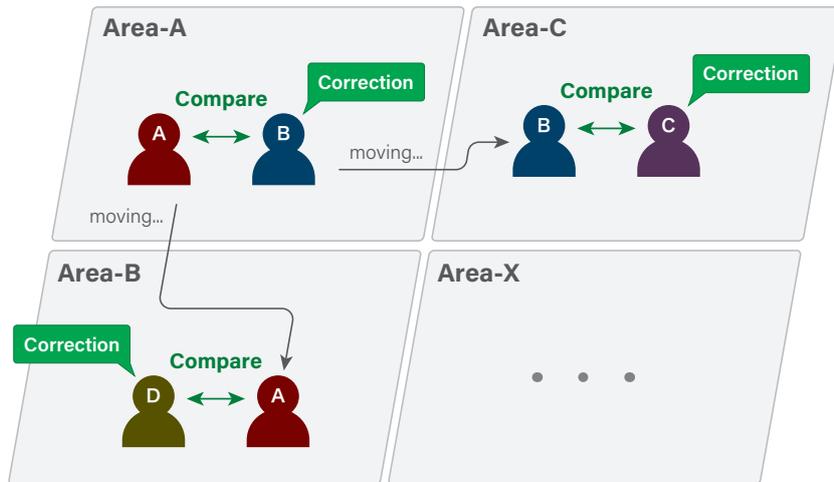


Figure 2.9. The concept of the relative correction method of measured street-lamp's illuminance at all areas.

In the following parts of this chapter, we describe the method for correcting OIL characteristic in Section 2.3.2, and IF characteristic in Section 2.3.3. Additionally, Section 2.3.4 provides the method for estimating illuminance of streetlamp using corrected illuminance data.

### 2.3.2 Correction of Obliquely Incident Light Characteristic

According to preliminary surveys, the illuminance transition when a user passes under a streetlamp on foot might drastically change affected by the OIL characteristic. Figure 2.10 shows the illuminance transition measured using two tablet devices (Nexus 7)<sup>\*4</sup>. It suggests these two devices have different OIL characteristic because of the difference of their transition shape (rise time and fall time).

Here, this section provides the method to correct the OIL characteristic of each devices based on the fitting of illuminance transition shape. Due to the unstable measurement conditions using hand-holding smartphone embedded sensors, we use illuminance data derived by averaging several samples for each streetlamp to reduce the dispersion between samples. In the following, we describe the procedure to correct illuminance transition shape:

1. Split up time series illuminance data into fragment data in each streetlamp. Then, convert splitted data to distance series data. The specific method will be described later.
2. Correct illuminance data for each measurement point using the angle of incident light calculated by smartphone holding angle and distance from the streetlamp position. Appendix B. provides the detailed explanation.
3. Remove high-frequency noise by applying weighted moving average (WMA) shown in Equation (2.1). The variable  $E_i$  represents illuminance value at index  $i$ . The weight of WMA has been empirically decided as  $w = 0.7$  in consideration of balancing between keeping transition shape and removing high-frequency noise.

$$E_i = w \cdot E_{i-1} + (1 - w) \cdot E_i \quad (2.1)$$

4. Normalize the transition shape for the range  $[0,1]$  to remove the effect IF characteristic.

---

<sup>\*4</sup> This transition is the average of several measurement samples targeted the same streetlamp. Furthermore, these data have been corrected using the smartphone holding angle and Lambert's cosine law, and then normalized.

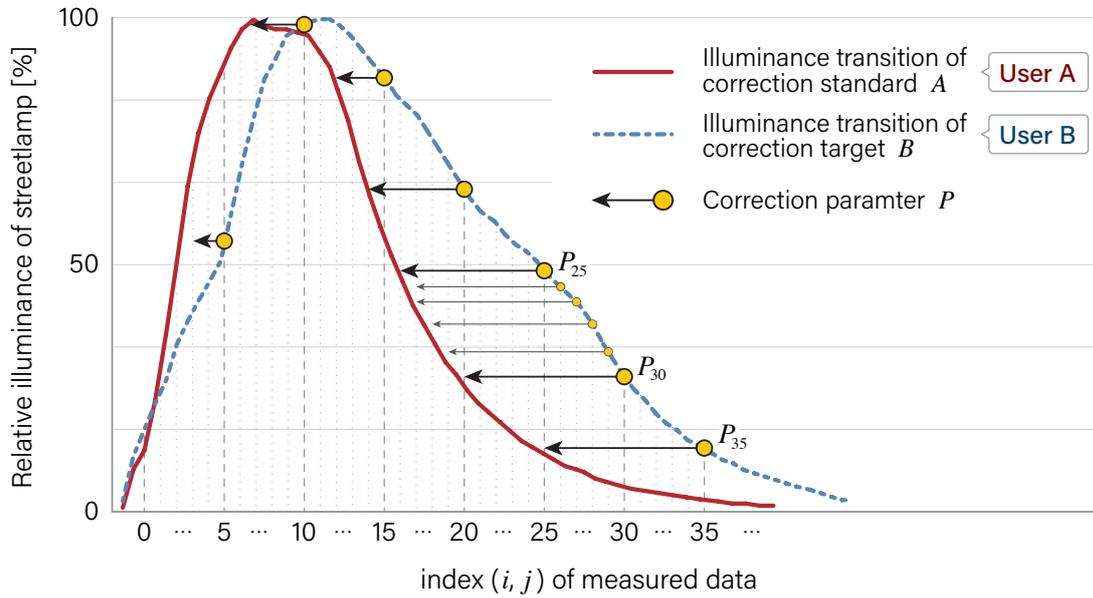


Figure 2.10. Correction parameters of an incident angle characteristics of illuminance.

5. Derive the average transition shape for each devices by averaging data in each measurement points.
6. Find the corresponding points between the transition shape of correction standard device (user A) and correction target device (user B) to calculate the differences of them as shown Figure 2.10. We call this differences as *correction parameter* for OIL characteristic. The specific procedure will be described later.
7. Correct illuminance transitions for subsequently data using correction parameter.

We describe the method to separate data for each streetlamp and convert it to the distance sequence in procedure 1. In this study, we assume an environment in which continuous measurements are made when the user walks in sidewalks at night, hence the illuminance data measured by smartphone embedded light sensors is continuous time series data. In the measurement on the alley, the illuminance changes near the streetlamp shows a pulse shape, and at other places

it keeps the lowest illuminance value (vicinity of 0 lx). Using this condition, it is possible to separate the data for each streetlamp by extracting the pulse. Also, the illuminance data per streetlamp can be converted into distance series data by deriving the distance between certain two points from the GPS information. In order to convert into distance series data, we employ the Hubeny formula which can calculate the distance between arbitrary two points from latitude and longitude data.

Next, we describe the method to search corresponding points between the transition shape of correction standard device (user A) and correction target device (user B), and derive the correction parameter  $P$  of the OIL characteristic. Figure 2.10 shows outline of the method. Illuminance transitions  $A$  and  $B$  are array data sampled at same intervals, and have two elements of the distance from just under the streetlight and the illuminance at that point. The parameter  $i, j$  represents the array index of illuminance transitions  $A$  and  $B$ . The correction parameter  $P$  of the OIL characteristic is also array of correction distances corresponding to the distance from the streetlamp, and  $P_j$  is an element of  $P$ . The sampling interval has been set to about 0.5 m in consideration of the walking speed of a user and the frequency of a light sensor.

1. Set  $i = 0, j = 0$  (As a side note,  $A_0 = 0, B_0 = 0$ ).
2. Increment  $i$  while  $A_i < B_j$  is true.
3. Set  $P_j = i - j$  (the correction parameter  $P$  at index  $j$ ). Then, increment  $j$ .
4. Repeat procedure 2 and 3. When both illuminance transitions reach to maximum ( $A_i = 1$  and  $B_j = 1$ ), proceed to next step.
5. Increment  $i$  while  $A_i > B_j$  is true.
6. Set  $P_j = i - j$ . Then, increment  $j$ .
7. Repeat procedure 5 and 6.

When there is a disturbance in the waveform such as two mountains exist, the above procedure can not be continued. In this case, the correction target  $B$  will be excluded because it is presumed that the illuminance transition has outliers.

### 2.3.3 Correction of Illuminance Following Characteristic

It is considered that the difference in the IF characteristic will be appeared as a difference in the absolute illuminance at each measurement point for the illuminance transition when a user walks under a streetlamp. However, since the transition shape is different as described above, it is impossible to compare absolute illuminance at all measurement points. In this section, we describe the correction method of the IF characteristic using the maximum illuminance as the representative measurement point. We correct this characteristic based on the illuminance data of the streetlamp measured a plurality of times, and correction parameter are calculated. The procedure is shown below.

1. Split up time series illuminance data into fragment data in each streetlamp. Then, convert splitted data to distance series data using the method described in Section 2.3.2.
2. Correct illuminance data for each measurement point using the angle of incident light calculated by smartphone holding angle and distance from the streetlamp position (see Appendix B.).
3. Remove high-frequency noise by applying WMA.
4. Apply the filtering which is based on transition shape. The specific method will be described later.
5. Compare the average maximum illuminance value of the correction standard and the correction target. Then, set this ratio as correction parameter of IF characteristic.
6. Correct illuminance transitions for subsequently data using correction parameter.

In the filtering by transition shape in procedure 4, the data much affected by external noise will be excluded based on comparison between average transition which is derived in procedure 5 of Section 2.3.2 and temporarily normalized transition. In an environment where a user measures while walking, there are a certain number of measured data for which the maximum illuminance has not

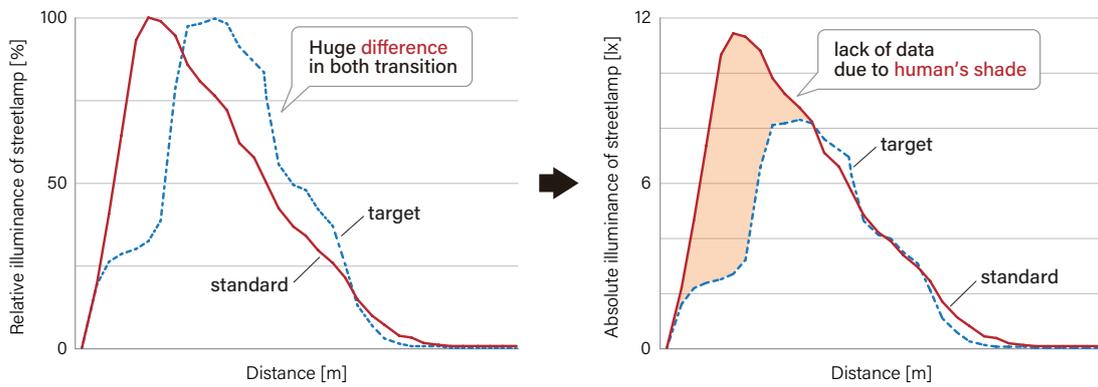


Figure 2.11. An adverse effect of a user's shadow.

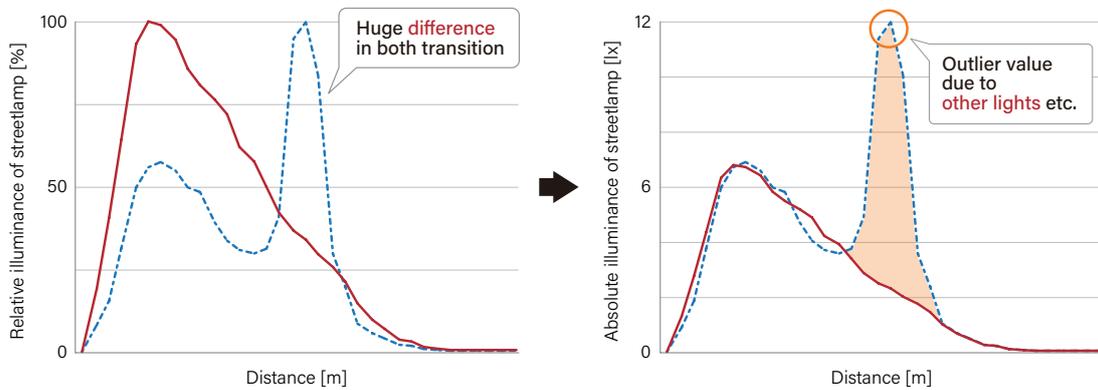


Figure 2.12. An adverse effect of an accidental light incident.

been correctly acquired due to some reasons. For example, the case of rays of light is intercepted by shadow of the user himself (Figure 2.11), or other light sources such as gate lights, automatic vending machines and head lights of vehicles is got mixed (Figure 2.12). Hence, as this filtering method, we have employed a square error at each measurement point of the transition is calculated and a threshold is set.

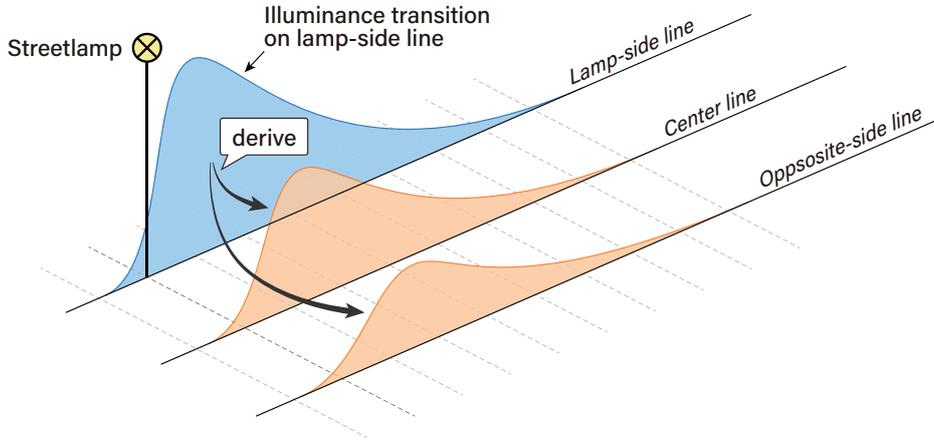


Figure 2.13. The inference of illuminances on the centerline of the road and the opposite line of the road.

### 2.3.4 Streetlamp Illuminance Estimating Method

Thus far sections, we have proposed the method for correcting two different characteristics of light sensors. Here, we proceed to the method for estimating illuminance of streetlamps according to JIS method using corrected illuminance transition data, which is a distance series data when a user passes under a streetlamp on foot.

First, the horizontal illuminance value at the measurement point just under the streetlamp is estimated based on OIL characteristic of correction standard device. In our condition, we have used a cosine to the 5th power ( $\cos^5$ ) as OIL characteristic of Nexus 7.

JIS method requires to derive average horizontal illuminance along the center, lamp-side, opposite-side lines of road shown in Figure 2.13. As for horizontal illuminance, the angle with the light source of streetlamp  $\theta_l$  only relies on the distance from streetlamp position. Horizontal illuminance along each line is derived with following procedures. Here, the value  $E^{\text{lamp}}$  and  $E^{\text{center}}$  represent the array of illuminance value measured on the lamp-side line and the center line respectively as shown in Figure 2.14, and  $i, j$  represent the array index of them. Also, the parameter  $d_i, d_j, x$  represent the distance between the streetlamp position and each measurement point,  $w$  represent the road width.

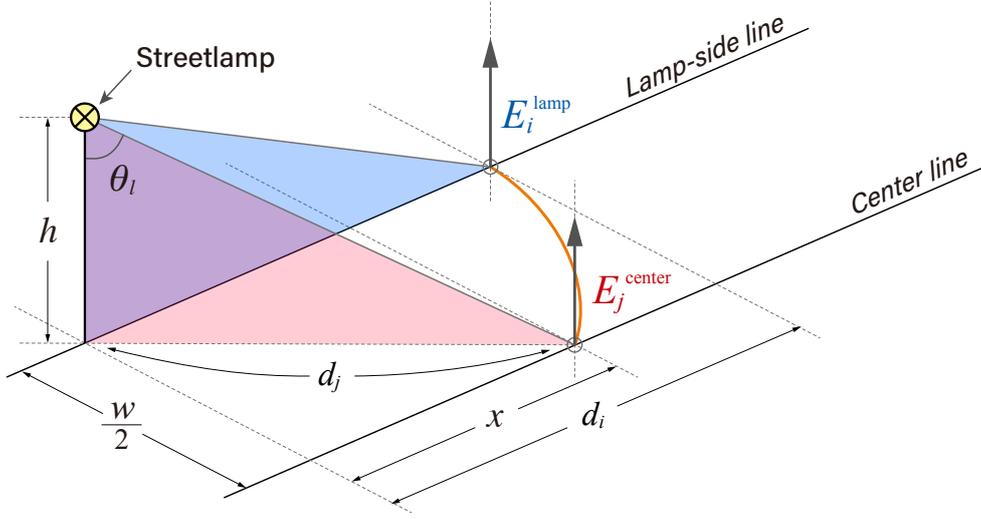


Figure 2.14. The inference of illuminances on the centerline of the road.

1. Set  $i = 0$  and  $j = 0$ .
2. Calculate  $d_j$  with Equation (2.2).

$$d_j = \sqrt{x^2 + \left(\frac{w}{2}\right)^2} \quad (2.2)$$

3. Increment  $i$  while  $d_i < d_j$  is true.

Then, derive the illuminance on the center line ( $E_j^{\text{center}}$ ) with Equation (2.3).

$$E_j^{\text{center}} = \begin{cases} E_i^{\text{lamp}} & (d_i = d_j) \\ E_{i-1}^{\text{lamp}} + R_{ij} \times E_{\text{diff}}^{\text{lamp}} & (d_i > d_j) \end{cases} \quad (2.3)$$

$$R_{ij} = \frac{d_j - d_{i-1}}{d_i - d_{i-1}} \quad (2.4)$$

$$E_{\text{diff}}^{\text{lamp}} = E_i^{\text{lamp}} - E_{i-1}^{\text{lamp}} \quad (2.5)$$

4. Repeat procedure 2 and 3.

Also, the illuminance on the opposite-side line ( $E^{\text{opposite}}$ ) can be derived with changing the road width parameter  $\frac{w}{2}$  to  $w$ .

As stated above, the illuminance transition data measured along each line can be derived, and the average horizontal illuminance by JIS method can be calculated with sampling these data at the same interval and averaging them.

### 2.3.5 Measures Against Propagation Error of Correction

We realize the accurate and comprehensive correction of illuminance measurement data with relative correction network (see Figure 2.15) generated by repeating procedures which regard corrected device as new correction standard device, as described in Section 2.3.1. However, there is a risk of propagation error which comes from the accumulation of the small error between the original correction standard by repeating correction procedures. As a side note, the relative correction network represents the relationship of a standard/target about the combination of each device, i.e, it has a traceability of correction process.

Therefore, we take measures using the labeling of correction counts. Figure 2.16 shows the overview of error propagation measures using labeling of correction counts when focusing on users A, B, and C shown in Figure 2.15. It is described in detail below. First, although the first correction is as described previous sections, the system records the correction counts (depth of the tree) for each device as a label when correcting. In the second and subsequent procedures, re-correction is performed by comparing this label. Finally, this label will be updated whenever each re-correction procedure. The parameter  $n$  and  $m$  represents labels of correction counts of users B and C respectively. The condition of re-correction is shown below.

- **case:**  $m + 1 < n$   
Re-correct user B regarding user C as a standard.
- **case:**  $m > n + 1$   
Re-correct user C regarding user B as a standard.
- **case:**  $m - 1 \leq n \leq m + 1$   
Not perform a re-correction procedure for both user B, C.

Repeating this procedure makes a depth of the tree shallow as shown in Figure 2.15, and can reduce propagation error as the result.

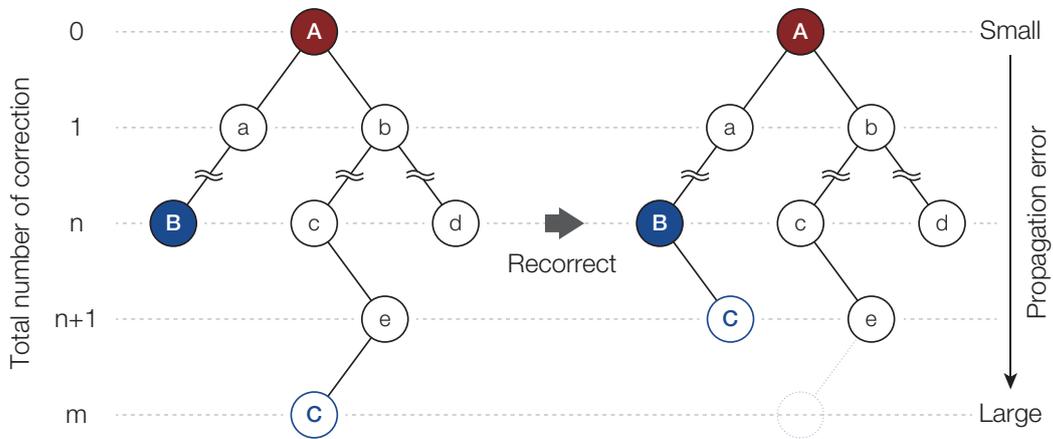
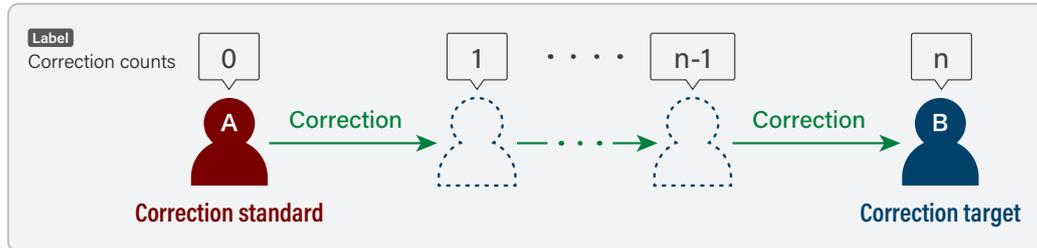


Figure 2.15. The history tree of the correction.

Correction on 1st time



Correction on 2nd or later time

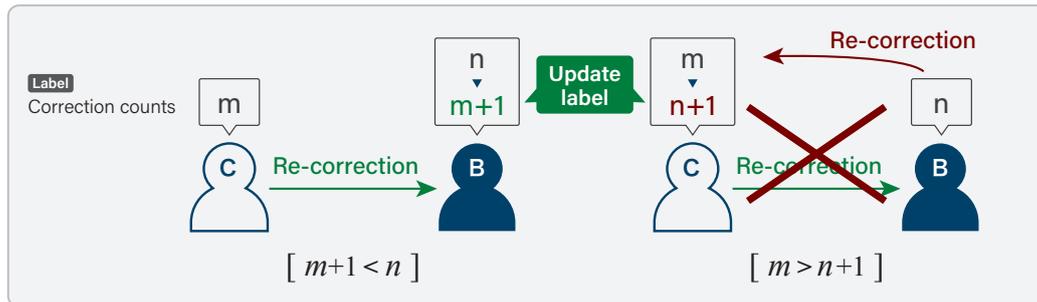


Figure 2.16. The reduction of the accumulative error of propagations by labeling of the number of times applying the correction method.

## 2.4 Experiments and Evaluation

### 2.4.1 Overview of Experiments and Evaluation Metrics

Here, we evaluate our method proposed in this chapter with real-world experiments. In the experiments, we have employed a route, that is located in Akashi, Hyogo, Japan, shown in Figure 2.17 as the experimental environment. This route is a typical Japanese alley, and has various kind of streetlamps: eight starter-type fluorescent lights, seven inverter-type fluorescent lights, and six LED lights (21 streetlamps in total). The circles drew in Figure 2.17 represent the position of each streetlamp, and the streetlamp No. is indicated next to each circle. The relationship of streetlamp No. and type of streetlamp is shown as follows:

- **Streetlamp No. 0–7** : Starter-type fluorescent lights (20W)
- **Streetlamp No. 8–14** : Inverter-type fluorescent lights (32W)
- **Streetlamp No. 15–20** : LED lights

As the experimental device, four Nexus 7 with different characteristics is used by individual users. Among the four devices, we use one as correction standard device, and three as correction target device. All users make measurements during the same experiment. We assume that users shall hold the device so that the screen of the device can be visually observed while walking at a natural speed during the experiment. Regarding the holding angle of the device, users could arbitrarily set without any specific conditions. We have selected two streetlamps for each its type (streetlamps No. 0–1, 8–9, 15–16) as for deriving correction parameter. Then, we calculate illuminance transition using data measured by correction standard/target device for each two streetlamp, and derive correction parameter with comparison of them.

In this paper, since the main purpose is to verify the validity of correction parameter extraction and illuminance transition correction. Hence, we conducted experiments in the environment which the effect of error propagation described in Section 2.3.5 does not occur by limiting the number of correction up to one.

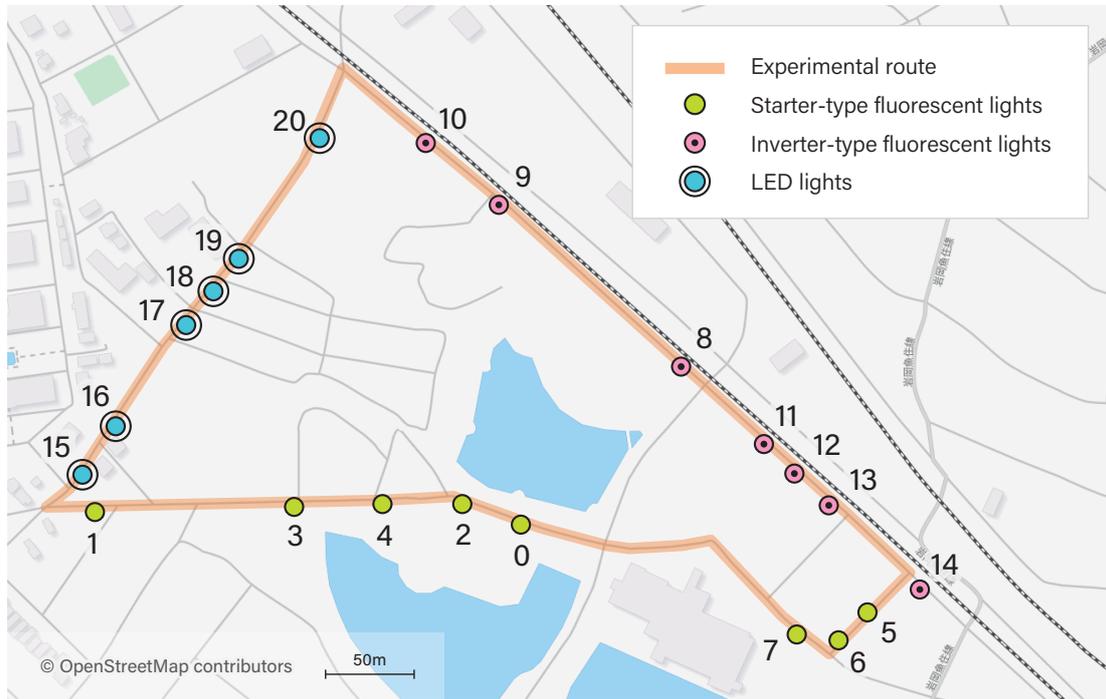


Figure 2.17. The evaluation field of the correction method.

To evaluate our method, we have conducted two experiments: the experiment of illuminance transition fitting, to confirm whether correction parameter of OIL characteristic and IF characteristic can correct the illuminance transition; the experiment of illuminance estimation using collected data, to confirm whether the accuracy of illuminance estimation will be improved by the correction. The evaluation method of each experiment will be described below.

### Experiment of Illuminance Transition Fitting

The illuminance data correction method is based on the shape fitting of illuminance transition. It suggests the gap in illuminance come of differences of sensor characteristics will appear as systematic offsets for each measurement point. hence, we evaluate the fitting accuracy of illuminance transition by comparing the following three data.

- Illuminance transition of correction standard  $W^{\text{standard}}$
- Illuminance transition before applying correction  $W^{\text{before}}$
- Illuminance transition after applying correction  $W^{\text{after}}$

To evaluate the fitting accuracy, we have employed the similarity of illuminance transition (SoIT) as a metric. SoIT ( $S^{\text{before}}$ ,  $S^{\text{after}}$ ) is sum of absolute error, and derived with Equation (2.6). The parameter  $W_d$  represents the illuminance value at distance  $d$  from streetlamp position, and  $D$  represents a maximum distance of the range in which light can reach.

$$S^{\{\text{before}|\text{after}\}} = \sum_{d=0}^D |W_d^{\{\text{before}|\text{after}\}} - W_d^{\text{standard}}| \quad (2.6)$$

### Experiment of Illuminance Estimation Using Corrected Data

We also evaluate the accuracy of streetlamps' illuminance estimation to confirm whether illuminance transition fitting method improves estimation accuracy.

The average horizontal illuminance according to JIS method is derived by the method proposed in Section 2.3.4 with corrected illuminance transition data. Then, we evaluate the accuracy by comparing each following estimated illuminance:

- Average horizontal illuminance  $\overline{E}_{\text{ans}}$  measured by a illuminometer according to JIS method (*ground truth data*).
- Average horizontal illuminance  $\overline{E}_{\text{raw}}$  measured by a light sensor without considering any sensor characteristics.
- Average horizontal illuminance  $\overline{E}_{\text{cos}}$  measured by a light sensor with correction based on general OIL characteristic characteristics.
- Average horizontal illuminance  $\overline{E}_{\text{cos}^5}$  measured by a light sensor with correction based on characteristics of the correction standard device.
- Average horizontal illuminance  $\overline{E}_{\text{pr}}$  measured by a light sensor with applying proposed method.

The accuracy of streetlamps' illuminance estimation is evaluated with absolute error  $\varepsilon_{\text{raw}}$ ,  $\varepsilon_{\text{cos}}$ ,  $\varepsilon_{\text{cos}^5}$ ,  $\varepsilon_{\text{pr}}$  derived with Equation (2.7).

$$\varepsilon_{\{\text{raw}|\text{cos}|\text{cos}^5|\text{pr}\}} = \overline{E}_{\{\text{raw}|\text{cos}|\text{cos}^5|\text{pr}\}} - \overline{E}_{\text{ans}} \quad (2.7)$$

## 2.4.2 Evaluation Result and Discussion

In the experiment, 1613 measurement samples were obtained for 21 streetlamps in the experimental field. This samples include 137 samples by the correction standard device and 1476 samples by the correction target device respectively. The evaluation and discussion for each experiment will be described below.

### Evaluation of Illuminance Transition Fitting

The evaluation result of illuminance transition fitting method is shown in Figure 2.18. SoIT  $S_{\text{before}}$ ,  $S_{\text{after}}$  is plotted for each streetlamp. We have found that illuminance transition fitting method increased SoIT for 90 % of streetlamps (19 out of 21), and SoIT is improved by 29 % on average. Furthermore, we have confirmed the great improvement can be found even inverter-type fluorescent lights and LED lights, which have high illuminance or high directivity negatively affecting the accuracy of estimation.

Here, we discuss the reason and future perspective about two samples (streetlamp No. 2, 12) which the SoIT have gotten worse.

Figure 2.19 shows the illuminance transitions in the case of streetlamp No. 2. Although it can be said that the transition shape is corrected correctly as expected, we found the error increased since whole transition shape has been moved along direction of travel. The reason of such movement is caused by the fork road exists near the streetlamp and the illuminance attenuates faster if a user passes from the east to the west. In this case, since the measurement point where the maximum illuminance is obtained moves to direction of the streetlamp position when averaging the data, and it is considered that the error increases by applying the correction parameter of the OIL characteristic.

Regarding streetlamp No. 12 as shown in Figure 2.20, we found the negative effect such as the unexpected amplification of illuminance due to the external light

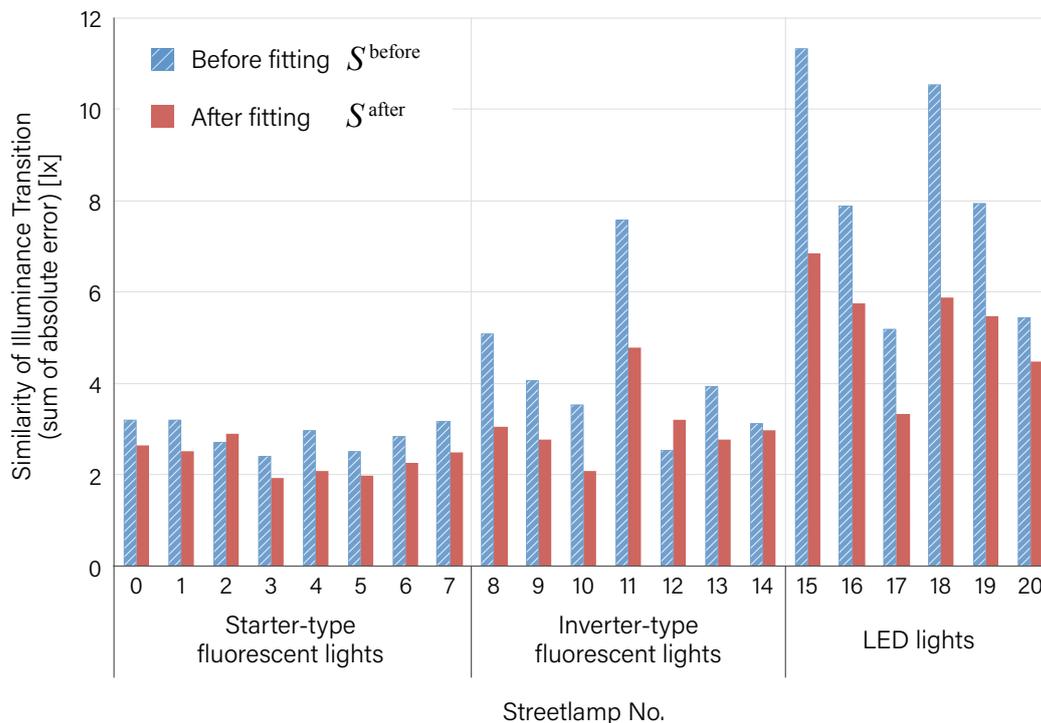


Figure 2.18. The result of the fitting method of illuminance transition.

source which has been installed in the opposite side of the road. This amplification was removed by the correction of the IF characteristic, and this correction causes the error near the streetlamp increased. Furthermore, this amplification changes the position where the maximum illuminance is obtained, it makes a problem similar to the case of streetlamp No. 2. While streetlamp No. 7 also have adjacent light source, this noise have not affect to the error because both of them have been installed at almost same place.

As the future work, we will investigate whether systematic differences will occur between straight road and refracted road, through deriving correction parameter in each condition. To deal with the external light source, we will consider the method to divide their illuminance by a technique such as a parameter estimation of mixture distribution.

Additionally, we have experimented in the condition of no covering the devices with any coatings such as protection sheet, it is not a realistic condition. In fact,

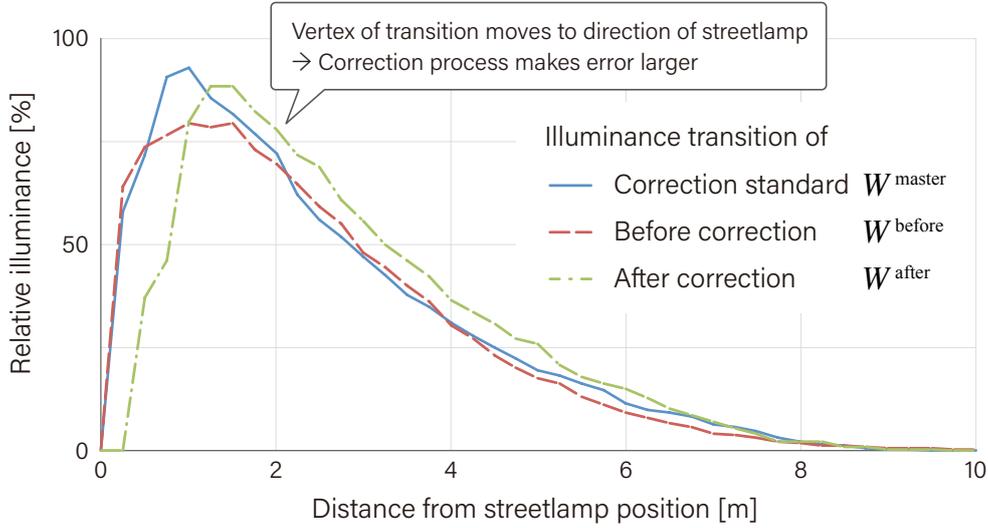


Figure 2.19. The comparison of illuminance transitions on streetlamp No. 2.

our method can be applied regardless of the existence of covering stuff because it is based on a relative correction. However, since such covering condition dynamically changes in daily life, we should invent the method of updating correction parameter regularly to keep correction accuracy, as a future work.

### Evaluation of Illuminance Estimation Using Corrected Data

The evaluation result of illuminance estimation method using corrected data is shown in Figure 2.21. The value  $\varepsilon_{\text{raw}}$ ,  $\varepsilon_{\text{cos}}$ ,  $\varepsilon_{\text{cos}^5}$ ,  $\varepsilon_{\text{pr}}$  is absolute illuminance estimation error mentioned in Section 2.4.1.

When sensor characteristics are not taken into consideration, the absolute error  $\varepsilon_{\text{raw}}$  is quite large (it may exceed 5 lx for inverter-type fluorescent lights and LED lights, and averaged absolute error  $\overline{\varepsilon_{\text{raw}}}$  is 2.33 lx). However, by applying our proposed method, the averaged absolute error  $\overline{\varepsilon_{\text{pr}}}$  in the estimation of streetlamp illuminance has been 0.54 lx. With our proposed method, the absolute error  $\varepsilon_{\text{pr}}$  has been improved around 75 % compared with using raw data ( $\varepsilon_{\text{raw}}$ ), and even improved around 20 % compared with applying our previous method ( $\varepsilon_{\text{cos}^5}$ ). In the following, we discuss the effectiveness of the proposed method for each type of streetlamp.

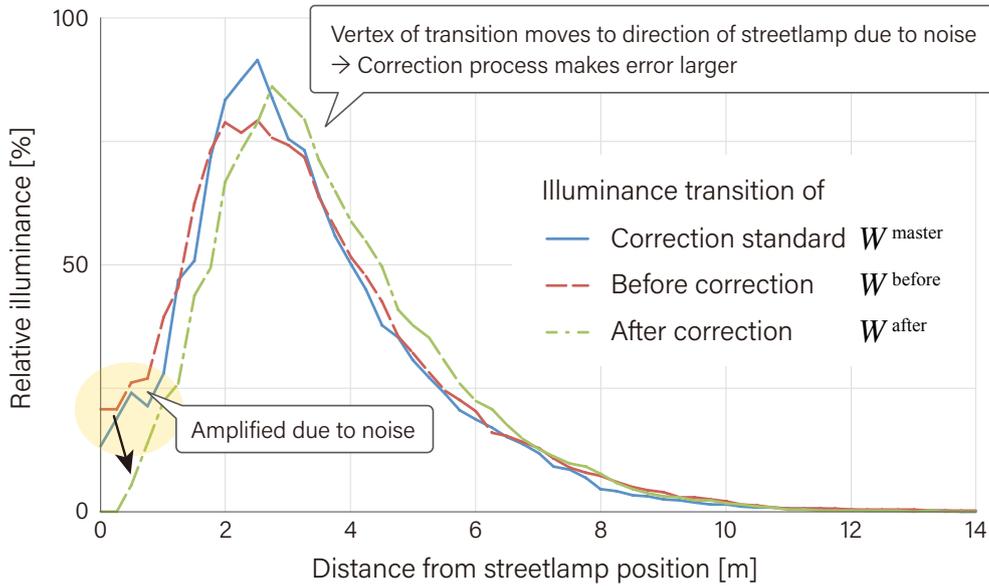


Figure 2.20. The comparison of illuminance transitions on streetlamp No. 12.

For starter-type fluorescent lights, it can be seen that the absolute error  $\varepsilon_{\cos}$ ,  $\varepsilon_{\cos^5}$ ,  $\varepsilon_{pr}$  is greatly reduced by applying correction of the smartphone holding angle. Furthermore, the averaged absolute error  $\overline{\varepsilon}_{pr}$  for all types of streetlamps was about 0.55 lx, which realized an error reduction of about 80 % compared with before applying the proposed method.

In streetlamps No. 5, 6, and 7, absolute errors have been increased by applying the proposed method. One of the reason is that these streetlamps are installed on sloping roads. The positional relationship between the device and the streetlamp, especially the incident angle of light, changes depending on the passing direction of the user. The proposed method assumes that streetlamps are installed at right angles to the road surface, therefore the error might increase when correcting with the smartphone holding angle. As for the streetlamp No. 7, the error increased due to the illuminance data was excessively corrected by the proposed method because the high illuminance light source is installed adjacent to the streetlamp.

Regarding inverter-type fluorescent lights, since this streetlamp has high illuminance, the error as a whole is large, however, the error can be reduced by applying the proposed method. Since the streetlamp No. 9 is installed on the

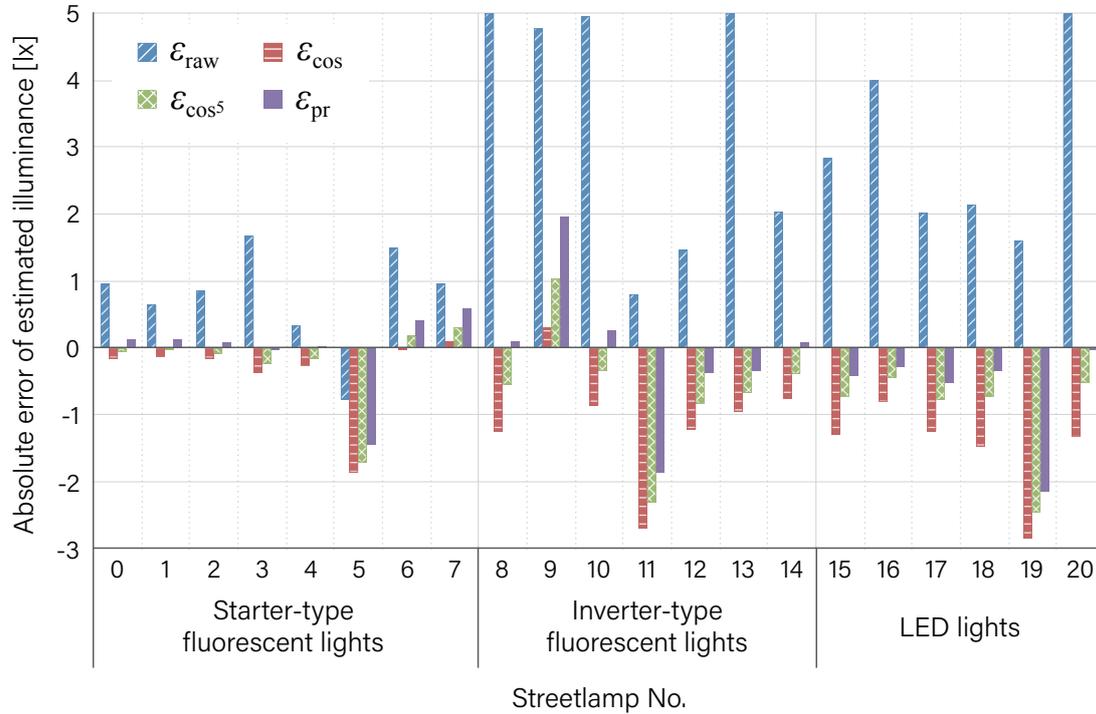


Figure 2.21. The result of illuminance estimation method with corrected data.

sloped road, it can be say the error is increasing as mentioned above. Also, the streetlamp No. 11 has a large error due to the influence of the high illuminance streetlamp with the same reason above.

As for LED lights, since it has the highest illuminance and directivity, and wide illuminated area, the measurement data is susceptible to noises. The result shows that proposed method can reduce the error even LED lights. However, considering the standard for safety assessment defined by JSSA, it can be said that the error of about 0.5 lx is still a large error.

From the above, in the condition assumed in this study, we have confirmed that the estimation accuracy is improved by applying proposed method regardless of the type of the streetlamp. As the future work, we should improve the robustness in various road conditions such as a sloped road and a forked road. For example, the inclination of the road can be estimated using a barometer, or the information provided by the government as open data also might be used.

## 2.5 Conclusion

In this chapter, we have focused on the objective context estimation method based on *human as a device carrier* concept. As a typical use case, we selected the topic of pedestrian safety, and aim to assess the safety level of sidewalks at night based on measurement of the illuminance of streetlamps by using a light sensor embedded a smartphone owned by general people. To realize accurate illuminance estimation using low-quality sensors, we proposed a method to relatively correct data measured by unknown devices using the data measured by devices whose sensor characteristics are known, and built the system named NightRoadScanner based on it. Through experiments in the real-world condition, we confirmed the proposed method reduced the estimation error for 90 % samples of streetlamps. Regarding the estimation accuracy, with our proposed method, the absolute error has been improved around 75 % compared with using raw data, and even improved around 20 % compared with applying our previous method. Finally, we have found our method can estimate illuminance of streetlamp with 0.54 lx averaged absolute error.

# 3 EmoTour: Urban Environment Analysis Using Human as a Sensor

In this chapter, as the sample case of sensing approach based on the *human as a sensor* concept, we describe a study on a estimation of users' psychological contexts during sightseeing based on the sensing of unconscious behavior of users.

## 3.1 Introduction - Background and Motivation

The great progress of sensing technology including the participatory mobile sensing explained in Chapter 2 have partially established the methodology to understand the complex urban environments. However, to provide more context-aware systems to users, the psychological context (e.g., emotional status) of users who are in urban spaces might also be an essential factor. For example, the geospatial emotion data can be used for improving user experience by inducing people. As the typical use case, we focused on tourism domain. In fact, the emotional status and satisfaction level of tourists are susceptible during sightseeing; hence, observing emotional feedbacks is useful for providing context-aware tourist guidance.

Recently, various approaches have been proposed to understand users' emotional status and satisfaction level in order to use them for consumer services. The most widely used approaches to collect satisfaction level of users are online user reviews and questionnaires [25, 26]. User reviews are also used in consumer services, such as the rating systems of TripAdvisor [27], Yelp [28], and Amazon [29]. However, keeping users motivated to regularly write reviews is difficult, especially with medium rating values, which leads to the risk of biased distribution

of review ratings. To provide reliable information for other users, it is necessary to collect quantitative data without user reviews. Though many studies have tried to estimate the emotional status of users with methods based on audiovisual data analysis [30–36], the accuracy of emotion recognition in outdoor places, such as tourist sights, tends to be worse due to the inclusion of environmental noise in audiovisual data [37–39]. In recent studies, the range of modalities has been expanding to physiological features (e.g., body movement, eye gaze) [40–48], and a fusion of them might help with even outdoor estimation [41, 49, 50].

Here, we have introduced the *human-in-the-loop sensing* framework, especially the *human as a sensor* concept, to the estimation of the user’s psychological contexts during sightseeing. In the following sections, we design a tourist emotion- and satisfaction-estimation system based on the sensing of their unconscious behavior, named *EmoTour* [51]. This system was developed in cooperation with Dmitrii Fedotov (Ulm University, Ulm, Germany), who performed the data analysis, emotion and satisfaction recognition models design and training, feature extraction in partially, cross-corpus and metamodelling, as a part of his doctoral dissertation.

## 3.2 Literature Review and Challenges

Due to the high demands of context-aware systems, especially in the tourist domain, there are many studies focused on environmental sensing to collect real-time information of tourist sights [4, 52]. Moreover, the recommendation algorithms based on tourists’ personal preferences have been discussed [53]. To provide intelligent recommendations in such systems, information about what people prefer what sightseeing place, i.e., psychological feedback from tourists, is required. However, the estimation of the psychological user context, which may be used for providing appropriate information based on the situation, has not yet been deeply tackled in spite of its importance.

In the following sections, we describe related literature that widely include other domains, then clarify the objective and challenges of our study.

### 3.2.1 Estimation of Emotional Status

Emotion recognition has been a hot topic for many research areas and for several years now due to the high demand for context-aware systems, such as *spoken-dialogue systems*. However, there are not many studies targeting the tourist domain yet.

In emotion recognition, audio- and/or visual-based approaches are popular fields in the field of dialogue systems and human–computer interaction [54]. In laboratory (indoor) conditions, existing audio-based emotion-recognition systems that use a deep neural network have achieved great performance [30,31]. Quirk et al. proposed an audio-based emotion-recognition system [32]. They built a dialogue system on mobile devices, and achieved around 60 % recall score for four affective dimensions. Tarnowski et al. proposed an approach based on facial movements [33]. They obtained good classification accuracy of 73 % for seven facial expressions. Moreover, they mentioned that head movements (orientation) could significantly affect extracting facial-expression features. Aiming at higher accuracies, bimodal emotion-recognition methods, combining audio and visual features, were also proposed [34,35], and they achieved better accuracy (e.g., 91 % for six emotion classes). However, in outdoor conditions, the accuracy of emotion recognition tends to be worse due to the inclusion of environmental noise in audiovisual data<sup>\*1</sup> [38,39]. Since tourist sights may have noisier environments, we should take such environmental conditions into account to estimate the emotional status of tourists.

To infer the emotion of a person, the unconscious behavior of humans may be a clue as well. Shapsough et al. described that emotions could be recognized by using typing behaviour on the smartphone [40]. This approach used a machine-learning technique and induced high accuracy on emotion recognition, yet it is not feasible to frequently ask users to type on their smartphone during a sightseeing tour. Resch et al. proposed an emotion-collecting system for urban planning called *Urban Emotions* [41]. The paper describes that wrist-type wearable devices and social media were used for emotion measurements. Since this approach relies on an assumption that posts on social media are written in situ, it has the

---

<sup>\*1</sup> According to “*Emotion Recognition in the Wild Challenge 2017 (EmotiW)*,” the accuracy of emotion recognition is up to 60 % for seven emotion classes [37].

problem of spatial coverage for collecting data. In recent large SNS, e.g., Twitter, Facebook, users cannot attach exact location data to a post with default settings, which makes UrbanEmotions difficult to collect comprehensive data. However, it also suggested that body movement can be used for recognizing emotion.

Moreover, recent studies in the field of emotion recognition focus on expanding the range of modalities and combining them. Ringeval et al. proposed to introduce physiological features in addition to audiovisual ones, and to build a multimodal system that relies on their combination [36]. As physiological features, an electrocardiogram (ECG) and electrodermal activity (EDA) were used. Physiological features provided lower performance and weaker correlation than audiovisual ones with continuous emotional labels, but helped to increase the overall performance of a multimodal system. Many studies also introduce physiological features: heart-related (ECG, heartbeat), skin- and blood-related (EDA, blood-pressure), brain-related (electroencephalography (EEG)), eye-related (eye gaze, pupil size), and movement-related (gestures, gyroscopic data) [41–48], and, in many cases, improvement of accuracy was observed, even in outdoor conditions [41, 49, 50].

### 3.2.2 Estimation of Satisfaction Level

In current consumer services, such as TripAdvisor [27], Yelp [28], and Amazon [29], online user reviews and questionnaires are still widely used to collect the satisfaction level of users. TripAdvisor [27] in particular uses a five-star rating system and comments from tourists as user reviews about sightseeing spots. To guarantee quantity and quality of voluntary reviews, it is essential to provide a motivation to contributors. However, keeping users motivated to regularly write reviews is difficult, especially with medium rating values, because many people do not like to post a review without any external incentives when they felt “*there’s nothing special*”. It means there is a risk of skewing the evaluation due to an imbalance of reviewers’ distribution.

Many studies also adopt questionnaire-based surveys for measuring tourist satisfaction [25, 26]. Fundamentally, several hundred samples (respondents) are required to produce reliable data. However, because the questionnaire-based method relies on the manual tasks of a human, it has problems in sustainability

and the spatial coverage of the survey. It also has the same risk as the method based on user reviews and ratings.

### 3.2.3 Objective and Challenges

Our objective was to determine the quantitative emotional status and satisfaction levels of users to design more intelligent and reliable guidance systems. However, through the investigation of current studies and services, we found several problems (e.g., biased reviews, spatial coverage of evaluation, accuracy of estimation) for applying existing techniques to real conditions of use.

From this background, the main challenge of our study was to establish a state-of-the-art method for estimating user emotion and satisfaction by fusing audiovisual data and various sensor data specialising in user behavior. In the following sections, we describe the design and implementation of a method for estimating the emotional status and satisfaction level of tourists, and provide deeper evaluation and discussion through real-world experiments.

## 3.3 Proposed Approach and Workflow

Our approach was designed for the tourist domain, where users are walking on a path through different sightseeing areas. To estimate the emotion and satisfaction of tourists during sightseeing, we focus on various actions that tourists unconsciously and naturally do during sightseeing. For example, tourists might approach a scenic place or work of art, stop, and gaze at it, potentially take selfie photos, or send video messages to their friends. The accumulation of these natural actions should be linked to the emotional status and satisfaction level that they feel there. Through a preliminary study, we have already confirmed that several actions (eye gaze and head/body movement) have a relationship with the emotion and satisfaction of tourists [55]. Hence, we propose an approach to observe such natural actions of tourists and estimate the tourist emotion/satisfaction while performing them.

Figure 3.1 shows the workflow of our whole system for collecting the data of tourist actions and labels. The overview of each step is as follows. And then, in the Section 3.3.1 and 3.3.2, we describe the details of the modalities and labels.

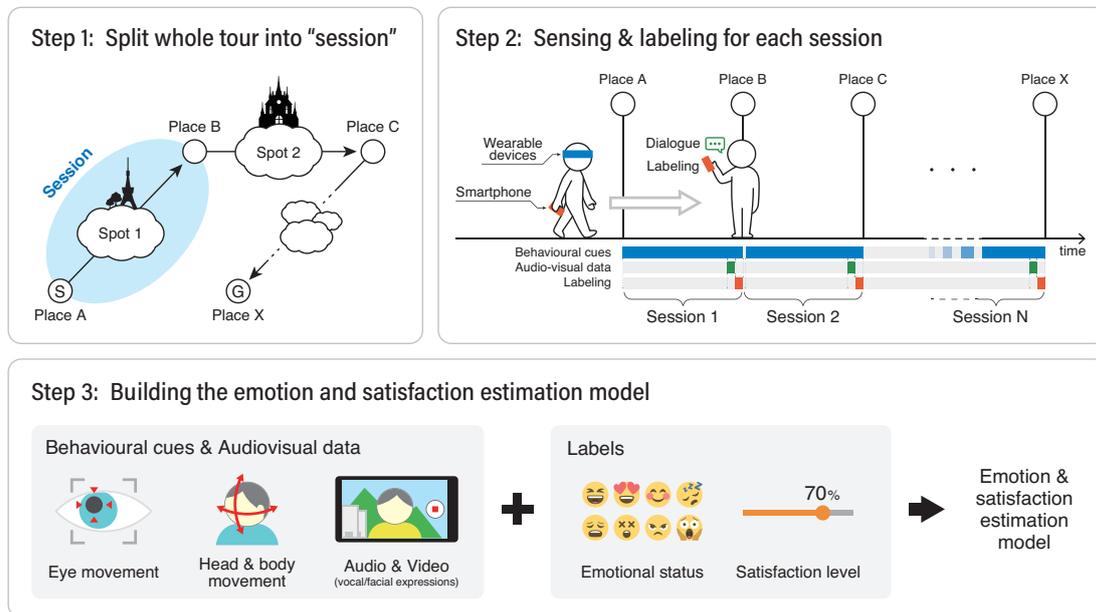


Figure 3.1. Workflow to estimate tourist emotional status and satisfaction level.

### Step 1 – Split the whole tour into sessions

Before starting sightseeing, we split the whole tour into small periods (sessions) that included at least one sight each. We assumed that a tourist typically requests guidance information for each sightseeing spot.

### Step 2 – Sensing and labeling

Tourists could freely visit sights while equipped with wearable devices that continuously recorded their behavior during the whole sightseeing. At the end of each session, they gave small amounts of feedback about the latest session by recording a selfie video. We assumed that recording a video serves as a means of interacting with dialogue systems or sending a video message to their friends. They also manually input their current emotional status and satisfaction level as a label. Then, they repeated the same procedure for each of the tour sessions.

### Step 3 – Building the estimating model

The tourist emotion- and satisfaction-estimation model was built based on tourist behavior, audiovisual data, and labels.

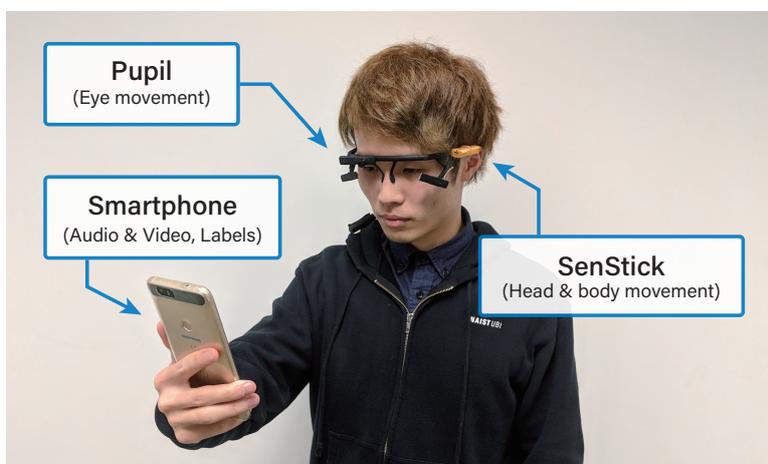


Figure 3.2. Devices for data collection during sightseeing: Pupil [56], SenStick [57], and smartphone.

### 3.3.1 Modalities

To perform an emotion estimation in the tourist domain, we used multimodal features: audiovisual data (vocal/facial expressions) and behavioral cues (eye and head/body movement data). Since tourists often take videos or photos, e.g., a selfie, audiovisual data could be used for our study. However, accuracy may have been low due to environmental issues in outdoor places, as mentioned in Section 3.2. Hence, we additionally used the features extracted from various tourist behaviors that happen unconsciously during sightseeing. The sense of sight is one of the most important sensory systems in sightseeing, and it can be tracked as eye movement using existing wearable devices and technologies. Moreover, due to the directivity on sensory systems (e.g., hearing, sight), head and body movements may be affected by them. Thus, we used head and body movements as features in addition to eye movement.

In our study, we used the three devices shown in Figure 3.2 to record features in real time: an Android smartphone (GPS-data, audiovisual data), a mobile eye-tracking headset Pupil with two 120 Hz eye cameras [56] (eye gaze, pupil features), and a sensor board SenStick [57] mounted on an ear of the eye-tracking device (accelerometer, gyroscope).

### 3.3.2 Labels

To represent the psychological context of tourists, we employed two types of metrics: emotional status and satisfaction level. We collected these data as labels by using the Android application shown in Figure 3.3. Tourists could manually enter the ratings of the session at the end of each session. The details of each metric are described as follows:

#### Emotional status

To represent the emotional status of tourists, we adopted the two-dimensional map defined on Russell’s circumplex space model<sup>\*2</sup> [58]. Figure 3.4 shows the representation of the emotional status. We divided this map into nine emotion categories and classified them into three emotion groups as follows:

**Positive** : Excited (0), Happy/Pleased (1), Calm/Relaxed (2)

**Neutral** : Neutral (3)

**Negative** : Sleepy/Tired (4), Bored/Depressed (5), Disappointed (6),  
Distressed/Frustrated (7), Afraid/Alarmed (8)

#### Satisfaction level

To represent the satisfaction level of tourists, we used the 7-Point Likert scale which the Japanese government (Ministry of Land, Infrastructure, Transport, and Tourism) uses as the official method. Tourists could choose their current satisfaction level between 0 (fully unsatisfied) and 6 (fully satisfied). A neutral satisfaction level is 3 and it should approximately represent the state of the participant at the beginning of the experiment.

---

<sup>\*2</sup> Russell’s circumplex space model is mainly employed for the time-continuous annotation of audiovisual databases that we used to build pre-trained models in Section 3.4.

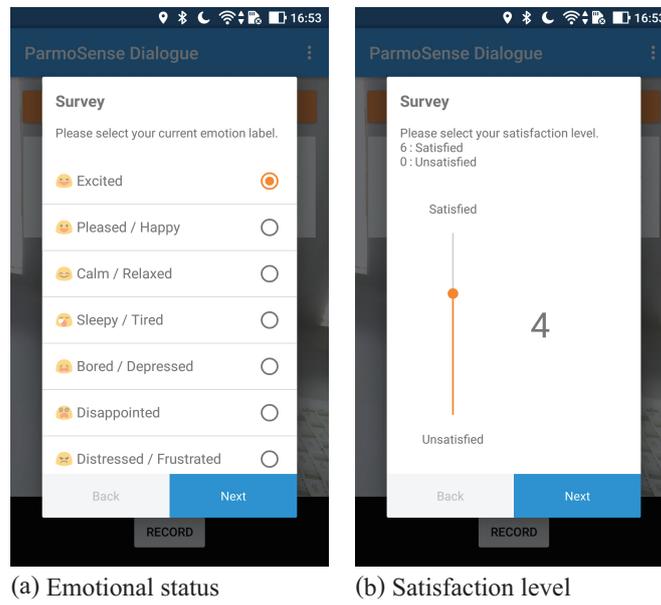


Figure 3.3. Smartphone application for collecting labels from tourists.

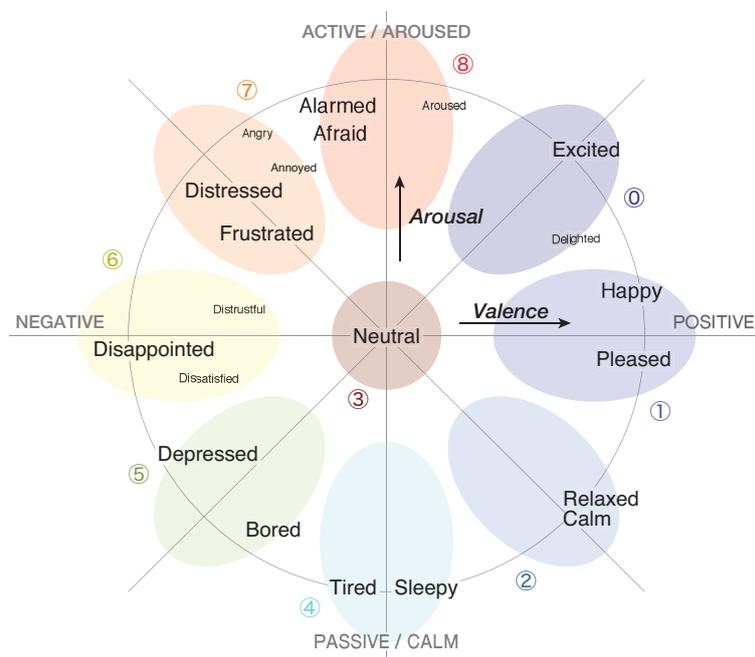


Figure 3.4. A two-dimensional emotion status model. This figure is derived from [58, 59].

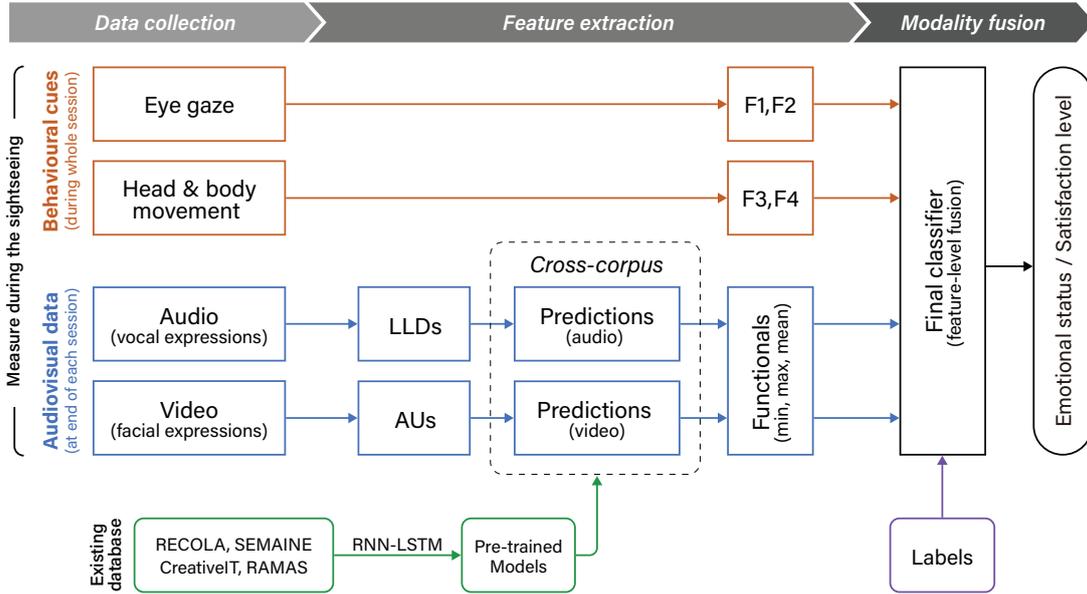


Figure 3.5. Scheme of tourist emotion and satisfaction estimation with modality fusion. LLDs: low-level descriptors of prosodic, spectral, cepstral, and voice quality, AUs: action units for describing facial expressions, F1: intensity of eye movement, F2: statistical features of eye movement, F3: head movement (head tilt), F4: body movement (footsteps), RNN-LSTM: recurrent neural network with long short-term memory.

### 3.4 Methodology of Tourist Emotion and Satisfaction Estimation

In this section, we describe the methodology of estimating emotion and satisfaction of tourists. Figure 3.5 depicts the scheme of our method, which consists of three stages. The first stage is *Data Collection* described in Section 3.3. The second stage is *Feature Extraction*, described in Section 3.4.1, where we preprocessed the collected raw data and extracted several features for each modality. The final stage is *Modality Fusion* described in Section 3.4.2, where several modalities were combined to build the final classifier of tourist emotion and satisfaction estimation.

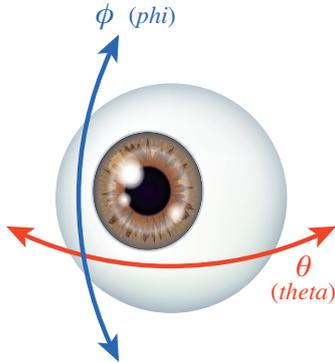


Figure 3.6. Spherical co-ordinates of pupil.

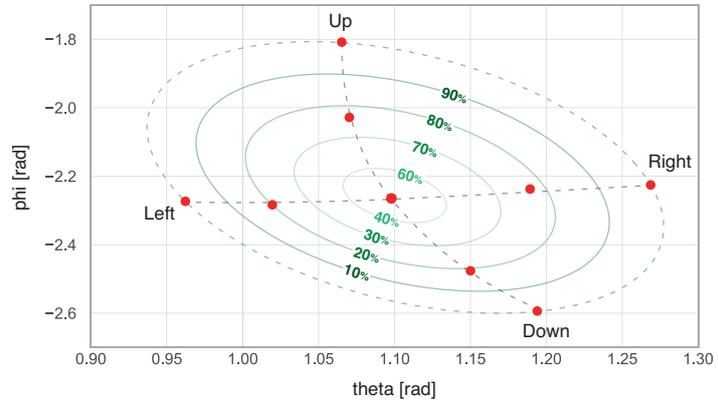


Figure 3.7. Threshold of eye movement (*theta* and *phi*).

### 3.4.1 Preprocessing and Feature Extraction

The raw data from each modality cannot be directly used to build tourist emotion and satisfaction estimation model. Hence, the methods of data preprocessing and feature extraction explained in Section 3.4.1 (behavioral cues) and Section 3.4.1 (audiovisual data) are applied to each modality.

#### Behavioral Cues—Eye-, Head-, and Body-Movement Features

**Eye-movement features** were extracted using the Pupil Labs eye-tracking headset [56]. We used *theta* and *phi* values, which represent a normal pupil as a 3D circle in spherical co-ordinates (Figure 3.6). Hence, we could only use two variables to describe the position of pupils and, thus, the eye gaze. Note that the raw values of eye-movement data differ across users and depend on the physical setting of camera and eye peculiarity. The eye-gaze data were analyzed using the following methodologies:

##### **F1: Intensity of eye movement**

Minimum and maximum values for *theta* and *phi* were calculated for each participant; eight thresholds (10–90 %, 10 % step, except 50 %) were set for the range [min, max] as shown in Figure 3.7, and then used to count the percentage of time outside each threshold per session. In total, 16 features were used.

## **F2: Statistical features of eye movement**

Average and standard deviation of  $\theta$  and  $\phi$  were calculated for a small window of recorded data and the values corresponding to the same session were averaged. The following window sizes were used: 1, 5, 10, 20, 60, 120, 180, and 240 s with the offset of  $\frac{1}{3}$  of the window size. In total, 64 features were used.

Then, **head and body movement features** were extracted using the inertial-sensor (accelerometer and gyroscope) values of the SenStick [57]. Both sensors have three axes: the X-axis, Y-axis, and Z-axis. Head- and body-movement data were analyzed using the following methodologies:

## **F3: Head movement (head tilt)**

As a head movement, head tilt was derived using gyroscope values. The average  $\mu$  and the standard deviation  $\sigma$  of the gyroscope values were calculated for each participant. Then, the upper/lower thresholds  $\psi$  were set with the following equations: Equation (3.1), Equation (3.2). The parameter  $a$  represents the axis of the gyroscope.

$$\psi_{upper,a} = \mu_a + 2\sigma_a \quad (3.1)$$

$$\psi_{lower,a} = \mu_a - 2\sigma_a \quad (3.2)$$

Finally, head tilt (looking up/down, right/left) was detected using threshold  $\psi$ . In our condition, the Y-axis indicates a looking-up/down motion, and the Z-axis indicates a looking-left/right motion. Since the duration of each session was different, we converted these data to several features: head tilt per second; and average and standard deviation of the time interval looking at each direction. In total, 23 features were used.

## **F4: Body movement (footsteps)**

Footsteps are analyzed with a method based on the approach proposed by Ying et al. [60]. First, the noises of accelerometer values were removed by applying a Butterworth filter with 5 Hz cutoff frequency. Then, high-frequency components were emphasised through the differential processing shown in Equation (3.3). The parameter  $x(n)$  represents the accelerometer

value at index  $n$ .

$$y(n) = \frac{1}{8} \{2x(n) + x(n-1) - x(n-3) - 2x(n-4)\} \quad (3.3)$$

Furthermore, the following integration process (Equation (3.4)) smoothed the accelerometer values, and small peaks of them were removed. In our condition,  $N$  was chosen to be 5 empirically<sup>\*3</sup>.

$$y(n) = \frac{1}{N} \{x(n - (N - 1)) + x(n - (N - 2)) + \dots + x(n)\} \quad (3.4)$$

Finally, footsteps were extracted by counting local maximum points. As features, we used footsteps per second, and average and standard deviation of a time interval for each step. In total, five features were used.

### Audiovisual Data—Vocal and Facial Expressions

**Audio features** (vocal expressions) were extracted with openSMILE software [61]. They consisted of 65 low-level descriptors (LLDs) of four different groups (prosodic, spectral, cepstral, and voice quality) and their first-order derivatives (130 features in total) used in ComParE challenges since 2013 [62]. The window size was set to 60 ms, and window step size was set to 10 ms, resulting in feature extraction on overlapping windows with a rate of 100 Hz.

As **Video features** (facial expressions), *action units (AUs)* were extracted with *OpenFace* [63, 64], an open-source toolkit. AUs describe specific movements of facial muscles in accordance with the *facial action coding system (FACS)* [65, 66], e.g., AU-1 means an action of “raising up the inner brow.” We used the following 17 AUs, which can be extracted using OpenFace: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, 45. The procedure of AU extraction is shown in Figure 3.8. First, the movie file is converted to a sequence of images, and facial landmarks are detected for each frame. Then, the face images are clipped out from the original frames because it expects that background objects and random people are partially captured in movie data in outdoor places. Finally, AUs are recognized by fusing several features taken from a clipped image for each frame of an original movie. In total, 18 AU-related features were extracted.

---

<sup>\*3</sup> Since the sensor position was different from the original method in our condition, we used a modified parameter.

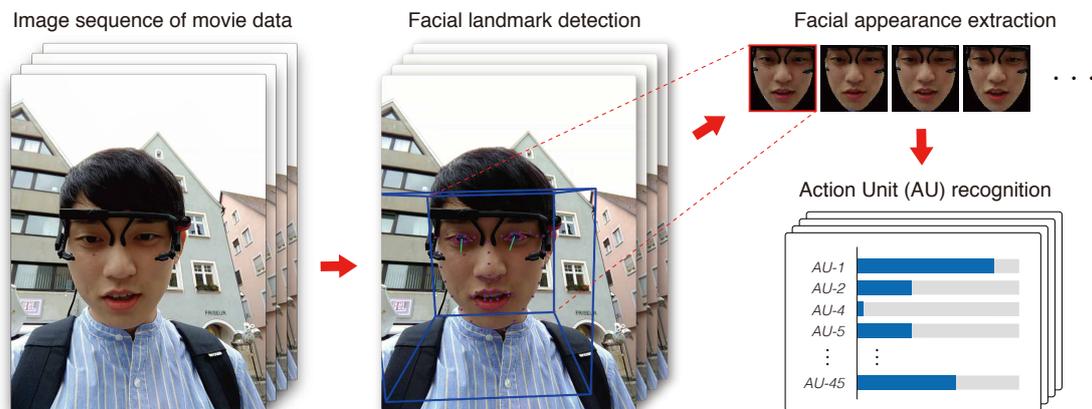


Figure 3.8. Video features (*Action Units*) extraction using *Openface* [63,64].

Due to the lack of data for training a decent audiovisual-based emotion recognition system, in this study we used models that are trained in advance by utilizing several existing corpora of emotionally rich interactions: the *RECOLA* (*Remote COLlaborative and Affective interactions*) database [36], *SEMAINE* (*Sustained Emotionally coloured MACHine-human Interaction using Nonverbal Expression*) database [67], *CreativeIT* database [68], and *RAMAS* (*The Russian Acted Multi-modal Affective Set*) database [69]. As corpora have different annotation rates, they were brought to the same data frequency to be able to share the same prediction models. The least frequency of 25 Hz, presented in RECOLA, was used for the remaining corpora.

The models are based on recurrent neural networks with long short-term memory (RNN-LSTM) and built in a way to consider the particular amount of context (7.6 s) in accordance to our previous study [70], as it shows better results. Networks were comprised of two hidden LSTM layers of 80 and 60 neurons with rectified linear unit (ReLU) activation function, respectively, each followed by a dropout layer with a probability of 0.3. The last layer has one neuron with linear activation function for regression tasks (databases: RECOLA, SEMAINE, CreativeIT) and six neurons with softmax activation function for the classification task (database: RAMAS). We used RMSProp [71] as an optimizer with a learning rate of 0.01. For regression tasks, we utilized a loss function based on

the concordance correlation coefficient that takes into account not only the correlation between two sets, but also the divergence, being not immune to biases. For the classification task, we used cross-categorical entropy.

After feeding the features extracted from the audiovisual data of our experiment to the trained model, for each time step we obtained the prediction in arousal and valence if we used the regression models, and probabilities of particular emotion if we used the classification models.

Models that are trained on the additional corpora cannot be directly used for emotional-status estimation in the context of our method due to the following reasons:

- Labels differ from those collected through our system in range and dimensions, i.e., they are on the arousal–valence scale instead of emotions for regression tasks, and an emotion set for a classification task does not match with ours.
- They are time-continuous, i.e., each value represents the emotional state for one frame of the audiovisual data, though we had one label per each session.

Taking these differences into account, predictions should be generalised and adapted to our method without losing valuable information. To achieve it, we took simple functionals (*min*, *max*, *mean*) from dimensional labels, i.e., arousal and valence as well as mean prediction scores for categorical emotions separately for each session, and merged them into feature vectors. Thus, we had high-level predictions from earlier trained models as features in our system.

For each modality, we used a simple feed-forward neural network with one hidden layer to make unimodal prediction from high-level features.

### 3.4.2 Modality Fusion

To build our final tourist emotion- and satisfaction-estimation system, we combined predictions based on our features, described in Section 3.4.1, on two levels: feature and decision. During the experiment, some problems with the devices and the data-collection process occurred that led to some data missing. The final

classifier should be robust and able to work with incomplete feature sets. On the decision-level fusion, it was achieved by applying linear models, where the final label is assigned, based on a linear combination of existing lower-level predictions. On the feature level, such feature sets were filled with zeros.

## 3.5 Experiments and Evaluation

### 3.5.1 Overview of Real-World Experiments

We conducted experiments in real-world conditions to evaluate the tourist emotion- and satisfaction-estimation method. As the experimental fields, we selected two tourist areas depicted in Figure 3.9 that have completely different conditions. The first one is the centre of Ulm, Germany. The sights in this area include particular buildings as well as walking routes with high tourist value (e.g., the Fisherman’s Quarter). The sights are surrounded by common city buildings and may be crowded depending on the time. The approximate length of the route is 1.5 km, divided into eight sessions. The second area is Nara Park, the historic outskirts of Nara, Japan. The route through the area includes many scenic and religious buildings (temples and shrines) that are located in nature, and has no distraction from the sights included in the sessions. The approximate length of the route is 2 km, divided into seven sessions.

Participants were asked to follow prepared routes and take as much time as they needed to see the sights. During the sessions, we recorded the data in real time according to Figure 3.1. At the end of each session, participants were asked to provide small feedback and labels for emotion and satisfaction level.

In total, we conducted our experiment with 22 participants, and collected 183 sessions’ data. The distribution of participants was the following: age range—22–31 years old (average age is 24.3); nationalities—12 Japanese, 10 Russian; gender—17 males, 5 females. In addition, 17 and 5 people went sightseeing to the tourist area located in Germany and Japan, respectively. Most of them were real tourists (e.g., short-term international students or new students, visitors), and hence they were not familiar with the experimental fields.

To collect the labels, the emotional status for each session during sightseeing could be measured only by the participants themselves. Due to the natural im-

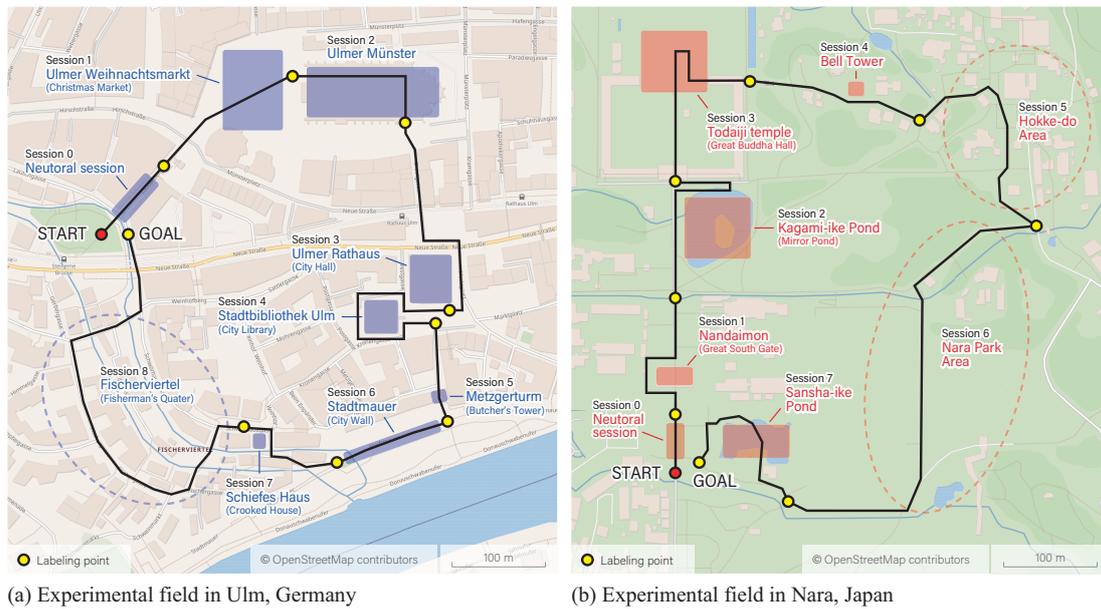


Figure 3.9. Experimental fields.

pression that sightseeing is something interesting, labels collected from tourists tend to be imbalanced. In our experiments, the distribution of labels was as shown in Figure 3.10, and the ratio of each emotion group was: positive: 71.0 %, neutral: 17.5 %, negative: 11.5 %. To manage this condition, we used *unweighted average recall (UAR)* as the performance metric. The same imbalance is presented for the satisfaction level as shown in Figure 3.11. Then, the relationship between emotion categories (groups) and satisfaction level is shown in Figure 3.12 (a)(b), and they suggest a moderate positive correlation ( $r = 0.644$ ), which proves the agreement between labels.

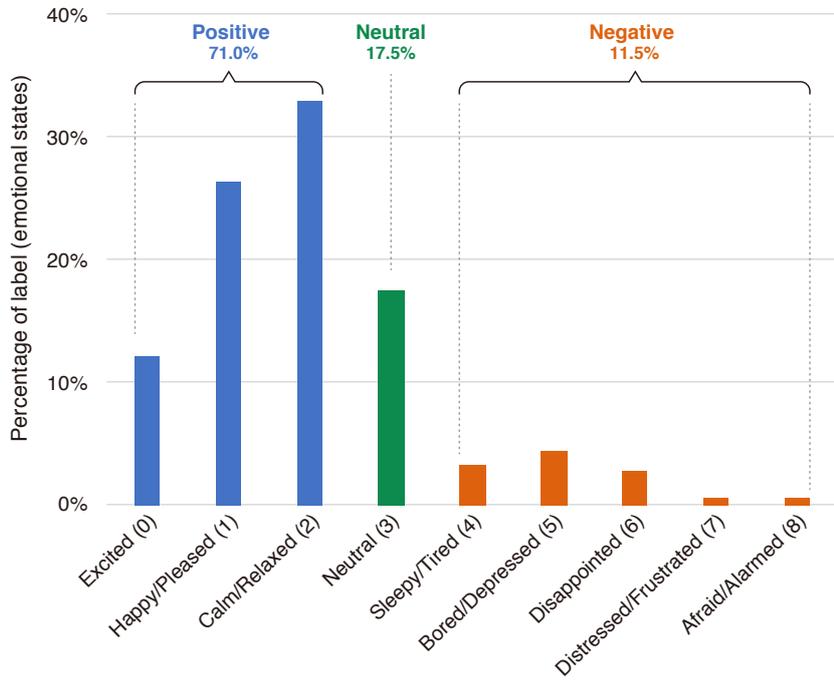


Figure 3.10. The distribution of labels (emotion).

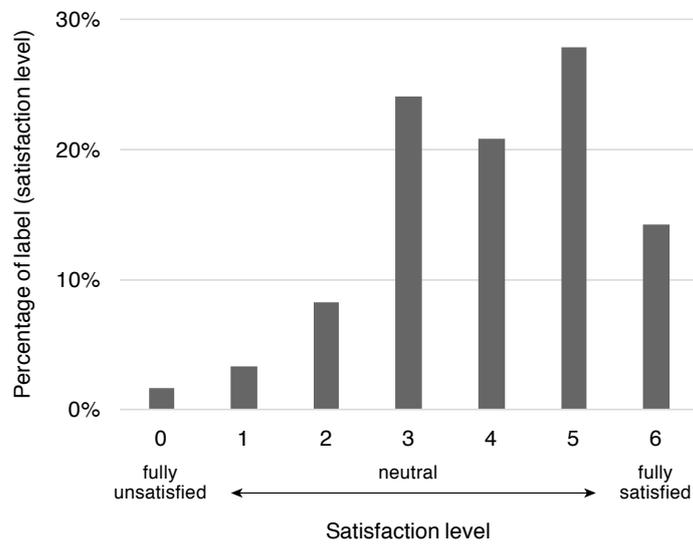


Figure 3.11. The distribution of labels (satisfaction).



Table 3.1. Performance of uni- and multimodal tourist emotion and satisfaction estimation <sup>\*a</sup>.

Modality	Emotion (UAR)		Satisfaction (MAE)	
	Avg.	SD	Avg.	SD
Eye movement (F1, F2)	0.432	0.073	1.124	0.178
Head/body movement (F3, F4)	0.428	0.070	1.187	0.170
<b>Behavioral cues (eye + head/body movement)</b>	<b>0.496</b>	0.130	1.171	0.188
Audio (vocal expressions)	0.410	0.069	1.124	0.154
Video (facial expressions)	0.404	0.092	1.101	0.155
<b>Audiovisual data (audio + video)</b>	0.431	0.098	1.108	0.165
<b>Feature-level fusion</b>	0.465	0.097	1.204	0.195
<b>Decision-level fusion</b>	0.485	0.098	<b>1.076</b>	0.134

<sup>\*a</sup> The best performances are highlighted with bold text.

### 3.5.2 Results

To evaluate our emotion- and satisfaction-estimation system, we conducted a series of reproducible experiments. To provide a fair comparison of different modalities or fusion methods, we fixed random seeds, resulting in consistent sample shuffling. The results for uni- and multimodal emotion and satisfaction estimation are presented in Table 3.1. We applied two different fusion approaches: feature-level fusion, where we built one model on merged feature vectors from corresponding modalities; and decision-level fusion, where we used a linear combination of prediction scores from unimodal models to set the final label.

We evaluated the performance of our emotion estimation system through a classification task of three emotion groups: Positive (0–2), Neutral (3), and Negative (4–8). Due to imbalanced label distribution, we used UAR as the performance metric. This ranged from 0 to 1 (the higher the better) and, for three-class classification problems, it had a chance level of 0.33. The results (Table 3.1) show that our system performed at up to 0.50 of a UAR score in three-class (all emotions) emotion-estimation tasks using head/body-movement features (F3, F4). Also,

the performance of 0.49 was shown with all features (decision-level fusion). According to statistical testing with Student’s  $t$ -test, there is no significant difference when using audiovisual data ( $p = 0.142$ ) or decision-level fusion ( $p = 0.755$ ). This suggests that there are implicit connections between emotional status and features, calculated for these modalities.

Then, to evaluate the performance of satisfaction estimation, we derived the *mean absolute error (MAE)* between estimated value and label. The MAE score can be any positive value, and 0 represents a perfect match. Table 3.1 shows that our system performed at a low MAE up to 1.08 in seven-level satisfaction-estimation tasks. The highest performance was shown with all features (decision-level fusion), and we confirmed there are significant differences compared with behavioral cues ( $p = 0.009$ ), audiovisual data ( $p = 0.076$ ) through  $t$ -test. It proves that combining behavioral cues and audiovisual data is useful to estimate the satisfaction level of tourists.

In our experiment, participants could be roughly divided into two almost equal groups by their nationality: Japanese (12 people) and Russian (10 people). It was proven by previous studies that different culture groups express emotions differently in ways and intensity [72–74]. To see whether the difference in cultural background affected the accuracy of estimation in our study as well, we conducted the same modeling procedures, described above, with two culturally differentiated groups of participants. Then, to evaluate the performance of emotion and satisfaction estimation, the UAR score and MAE were derived, respectively. Also, we have conducted statistical testing of UAR score and MAE between Japanese and Russian.

The results are shown in Table 3.2, and they suggest that the impact of nationalities or cultural difference exists on several features for each estimation tasks. For behavioral cues, we have found combining features of eye movement and head/body movement contributes drastic improvement of performance for Russian (but, not for Japanese) with emotion-estimation task. Then, we confirmed that their behavioral cues showed the highest UAR score of 0.58. For audiovisual data, we have also confirmed significant differences between Japanese and Russian. It is expected to be caused by a difference that how to express their emotion or satisfaction into voice or face.

These results suggest that we need to take the effects of nationalities or cultural differences into account in order to generalize the model of our system. As future work, we will expand the nationalities of tourists, and build a general model using them in consideration of their cultural background. In addition, we aim to investigate the effects of other attributes of tourists, such as gender and age.

## **3.6 Discussion and Limitations**

### **3.6.1 Feasibility of Proposed System**

The results from Section 3.5.2 prove that tourists' emotion and satisfaction estimation can be extracted by our proposed system to a certain degree. In our study, modality fusion showed a better result compared to unimodal systems for estimating both emotional status and satisfaction level. Especially combining behavioral features at a feature-level often improved the performance of emotion estimation, compared to eye- and head-based feature alone. The possible reason for this is that, according to the process of data collection and human movements, eye-gaze and head movement are connected to each other: a human moves them both while exploring an environment, usually replacing a significant eye movement with a slight head movement. The combination of these modalities at a feature level allows the system to simultaneously utilize information from both sources, which is not possible to do at a decision level.

### **3.6.2 Imbalance of Labels**

Studies related to emotion estimation often suffer from the subjectivity of labels. In our study, we had an exceptional case of subjectivity, as emotion and satisfaction can be measured only by the participants themselves and not by any third parties, such as an annotator. An additional limitation was brought by the domain of our research—tourism. As the main idea was to measure people's first impression, they could not participate twice in the same experiment, and should not be familiar with the experimental field. This means that we could not ask local citizens to participate in an experiment, which constrained the range of potential candidates to a very narrow group. These conditions resulted in a

Table 3.2. Performance of tourist emotion and satisfaction estimation (by nationality of participants)<sup>\*a</sup>.

Modality	Emotion (UAR)						Satisfaction (MAE)					
	Japanese		Russian		<i>p</i> -value ( <i>t</i> -test)		Japanese		Russian		<i>p</i> -value ( <i>t</i> -test)	
	Avg.	SD	Avg.	SD			Avg.	SD	Avg.	SD		
Eye movement (F1, F2)	0.438	0.086	0.426	0.061	0.742		1.001	0.142	1.248	0.114	0.006 *	
Head/body movement (F3, F4)	0.417	0.082	0.438	0.056	0.573		1.198	0.228	1.176	0.093	0.807	
<b>behavioral cues (eye + head/body movement)</b>	0.415	0.067	<b>0.576</b>	0.129	0.011 *		1.093	0.237	1.249	0.071	0.119	
Audio (vocal expressions)	0.447	0.069	0.372	0.048	0.010 *		1.032	0.098	1.217	0.146	0.020 *	
Video (facial expressions)	0.463	0.098	0.346	0.027	0.008 *		1.019	0.110	1.184	0.152	0.039 *	
<b>Audiovisual data (audio + video)</b>	0.445	0.092	0.417	0.106	0.587		1.014	0.106	1.201	0.164	0.030 *	
<b>Feature-level fusion</b>	0.423	0.048	0.507	0.117	0.086 *		1.124	0.239	1.285	0.095	0.147	
<b>Decision-level fusion</b>	<b>0.473</b>	0.064	0.496	0.125	0.566		<b>0.995</b>	0.103	<b>1.157</b>	0.112	0.010 *	

<sup>\*a</sup>The best performances are highlighted with bold text. \* There is significant difference ( $p < 0.10$ ).

data shortage, complicating the model training stage, affecting the general performance and statistical stability of results.

Natural perception of sights as something interesting and the general conditions of an experiment led to a great imbalance of emotional labels. Some emotion groups are predefined to be almost empty a priori because these emotions can be caused by sightseeing, e.g., distressed or afraid. Even after dividing emotions into three groups (positive, negative, and neutral) unequally, we had 71 % of positive samples. Many of these problems and limitations can be partially overcome by increasing the amount of data, which would make the system more stable and robust.

### 3.6.3 Limitation of Data Sources

To realize the proposed system, we need to collect several data sources, such as eye-gaze data, head motions, and selfie videos. However, there is a limitation in collecting such data in real-life conditions. Although eye-tracking devices are becoming smaller and cheaper year by year, it will take more time for them to be commonly and frequently used. However, head-motions can be measured because JINS MEME [75] with electrooculography (EOG) and a six-axis inertial measurement unit (IMU) are already being sold on the market, and many people are using them in their daily life. In case of using selfie videos, we must take it into account that the emotional status on such videos may be exaggerated. One of the most possible shifts may be done from natural emotions to acted ones. If so, it is possible as future work to utilize existing databases for emotion recognition to improve the performance of the system.

The aim of our study, as our first attempt, was to reveal what kind of data are needed for estimating the emotional status and satisfaction level of tourists. Hence, we employed all kinds of conceivable modalities in our work, but we do not assert that all the modalities are required. Through the experiments, we have confirmed that various modalities and their fusion can be used for estimating emotional status and satisfaction level. We also found that there is no notable difference in the estimation performance at the unimodal level; however, performance can improve by combining them in several ways. This suggests that our proposed method does not rely on a specific modality or combination. In the

current situation, not many tourists take a selfie *video* that can be used as one of the data sources in our system. However, even without videos, our proposed method can perform at a certain level. Of course, if the selfie video becomes common, like selfie photos, performance can be improved.

### 3.6.4 Future Perspectives

The result of this chapter provides a baseline performance for estimating the emotional status and satisfaction level of tourists. Several ways can be considered to improve performance. One is to widely explore other available modalities and their combinations. For example, foot motion and direction might be used for estimating the degree of tourist interest during sightseeing. Another is to analyze the transition of emotional status and satisfaction level of tourists during a session. Since emotional status and satisfaction level can drastically change even in the same session, this transition process might also be a valuable clue to estimate the tourists' status at the end of the session.

## 3.7 Conclusion

In this chapter, we have focused on the subjective context estimation method based on *human as a sensor* concept. As a typical use case, we selected the tourist domain, and aimed to estimate the emotional status and satisfaction level of tourists during a sightseeing tour based on their unconscious and natural actions, e.g., selfie videos, body movements, etc. We proposed a tourist emotion- and satisfaction-estimation method by fusing several modalities. To build the model, four kinds of modalities were employed: behavioral cues (eye and head/body movement) and audiovisual data (vocal/facial expressions). Through experiments in the-real-world with 22 tourists, we achieved up to 0.48 of UAR score in the three-class emotion estimation task, and up to 1.11 of MAE in the seven-level satisfaction estimation task. In addition, we found that effective features used for emotion and satisfaction estimation are different among tourists with a different cultural background.

# 4 ParmoSense: A Participatory Mobile Sensing Platform

In this chapter, we come down to the problem how to keep sustainability of *Human-in-the-loop sensing* in the real-world, and describe a study on a participatory mobile sensing platform which the general people can organize/contribute the *Human-in-the-loop sensing* task.

## 4.1 Introduction - Background and Motivation

In Chapter 2 and Chapter 3, we described a positive potential of the *Human-in-the-loop sensing* framework through the design and evaluation of two use cases. However, due to its mechanism established on *cooperation* in which depends on the voluntarism of general people, the sustainability of the *Human-in-the-loop sensing* framework is a critical challenge in real-world operation. In this study, we focus on the mutual linkage with the local community, e.g., civic tech community, where ecosystems are already formed can be considered. To continuously operate participatory sensing systems in such communities for broad urban environment analysis, a platform that can be easily and quickly customized by organizers to perform a variety of sensing tasks, and that is easy to set-up and run on the participants' smartphones, is essential.

Through the investigation of existing participatory sensing platforms [76–81], we found two main challenges in using these platforms. The first challenge is *limited support of essential functions*. Existing platforms tend to focus on specific sensing purposes, e.g., urban transport data sensing, and therefore support limited functions. Because the purpose of sensing differs among organizers of urban sensing, the necessary functions, i.e., sensing function, incentive mechanism, task

request control, and data processing method, will also differ depending on the purposes. Thus, in the ideal participatory sensing platform, flexibility to adapt the platform to perform sensing for various purposes is mandatory. Moreover, motivating participants is an important aspect of participatory sensing since participatory sensing relies on voluntary participation of ordinary people [82], but we found that these platforms do not implement it enough. The second challenge is *difficulty of system construction and operation*. The platforms require a high level of technical skill for users. For example, some platforms require programming skills for organizers and data processing skills for participants. In order to open the door of participatory sensing to non-technical users, it is necessary to ensure that participatory sensing systems can be easily constructed and operated by both organizers and participants.

Here, we have designed and built a novel participatory sensing platform named *ParmoSense*, for easily and flexibly collecting urban environmental information for various purposes by overcoming the challenges mentioned above.

## 4.2 Literature Review and Challenges

This section is devoted to clarifying what kind of functions are necessary for participatory sensing systems and what kind of skills are required for users (organizers and participants) of participatory sensing systems. We first organized the necessary functions into three categories: sensing functions, motivating functions and processing functions. Then, we investigated the functions implemented in existing participatory sensing platforms [76–81]. The results are summarized in Table 4.1. The skills required by organizers and participants in existing participatory sensing platforms are shown in Table 4.2.

### 4.2.1 Functions Essential in Participatory Sensing Platform

#### 1) Sensing functions

Sensing is an essential part of participatory sensing systems. We define sensing functions as those that allow the organizer to specify what kind of sensors to

Table 4.1. Overview of supported functions

Platforms	Functions							
	Sensing functions		Motivating functions			Processing functions		
	F1	F2	F3	F4	F5	F6	F7	F8
	Implicit-sensing	Explicit-sensing	Request	Reward	Feedback	Editing	Browsing	Export
AWARE [76]	✓	△ <sup>*b</sup>			△ <sup>*g</sup>	✓	✓	✓
Medusa [77]	✓	△ <sup>*d</sup>	△ <sup>*e</sup>	△ <sup>*f</sup>			✓	
Funf [78]	✓	△ <sup>*d</sup>				✓	✓	✓
MinaQn [79]	△ <sup>*a</sup>	△ <sup>*b</sup>	△ <sup>*e</sup>		✓	✓	✓	✓
Ohmage [80]	△ <sup>*c</sup>	✓			△ <sup>*g</sup>	✓	✓	✓
OpenDataKit [81]	△ <sup>*a</sup>	✓				✓	✓	✓
ParmoSense	✓	✓	✓	✓	✓	✓	✓	✓

<sup>\*a</sup> Location (GPS) only supported. <sup>\*b</sup> Media upload (photo, sound, etc.) limited. <sup>\*c</sup> Raw data collection unsupported.

<sup>\*d</sup> Questionnaire unsupported. <sup>\*e</sup> Static request only supported. <sup>\*f</sup> Monetary incentives only supported.

<sup>\*g</sup> Feedback of data collected by oneself only supported.

use, and how to collect sensing data from the urban environment. There are two different ways of sensing: implicit (F1) and explicit (F2) sensing.

**Implicit-sensing (F1):**

Uses sensors embedded in mobile devices. It is mainly used for collecting urban environmental data without actively involving the participant, i.e., implicit sensing.

**Explicit-sensing (F2):**

Used for collecting data generated by human behavior, e.g., photos, voice, and questionnaires. It is used for collecting urban environmental data through directly involving participants, i.e., explicitly, and locally.

AWARE [76] provides a platform for both implicit and explicit sensing. For implicit-sensing, the organizer can choose which smartphone sensors to use from a web UI. They can also configure the detailed settings such as sensing interval. For explicit-sensing, AWARE allows the organizer to distribute questionnaires manually.

Ohmage [80] supports explicit-sensing, by allowing participants to post reports by themselves. Several report formats are accepted such as single/multiple selections, free text, multimedia (e.g., photo, sound) and so on. Ohmage also supplements collected data through implicit-sensing. Specifically, it can be used to record the transportation status (e.g., still, walk, run) of a participant.

The other conventional platforms however, tend to focus more on either implicit- or explicit-sensing (as shown in Table 4.1), and provide limited functionality for the other form of sensing. As mentioned before, the two sensing methods have many differences such as data type that can be collected, and the data's features. Thus, with these platforms, it is difficult to supplement the collected data due to severe restrictions in sensing functions.

In this study, in order to realize a flexible sensing platform, we aim to provide functions for both sensing methods comprehensively and support the organizers' ability to easily choose and combine functions.

## 2) Motivating functions

Since participatory sensing relies on voluntary participation of ordinary people, it is essential to not only focus on acquiring users, but also on motivating them to continue participating in the sensing tasks, i.e., to support user retention and activation [83]. Motivating functions allow the organizer to define the methods for motivating participants. Some of the conventional platforms use outside stimuli to support participant motivation and engagement. We classify these outside stimuli as follows:

### **Request (F3):**

This is urging behavior by explicitly requesting participation. There are many methods of sending requests. The most common method is the static/dynamic request, which includes providing a task list, issuing notifications and so on. Other methods such as Audition [77] and Reverse-auction [84,85], which purposely restrict the rights of contribution and make participants scramble to contribute, have also been used.

### **Reward (F4):**

In this method, participants are compensated for their contribution through monetary incentives and non-monetary incentives [86,87].

### **Feedback (F5):**

This method urges behavior by providing feedback such as a visualization of participants' contribution on a map, graph or timeline. Sometimes the contributions of other participants are also included in the visualizations, and sometimes they are excluded.

MinaQn [79] uses a recruitment mechanism to grant participants the right to participate in urban planning (contribution to society) as non-monetary incentive. In addition, the platform also visualizes a summary of the participant's contribution to increase their willingness to continue contributing.

Medusa [77] is a platform which utilizes monetary incentives effectively. Medusa can acquire participants by using recruitment, which provides money as compensation, and through audition. Furthermore, to retain participants, Medusa adopts the concept of reverse incentive (obligation/responsibility of executing

tasks), where workers pay to organizers for the privilege of performing the task. This can help prevent participants from quitting the system in the middle of sensing tasks.

In these conventional platforms, functions to motivate participants have a number of shortcomings. For example, with Request functions (F3), in order to increase the number of successful requests, it is necessary to consider the notification timing and the target participants, but this is not supported. Similarly, with Reward functions (F4), we need to consider not only monetary incentives, but also *gamification*, which is a non-monetary incentive mechanism that gives *experience* (a kind of fun) as compensation for participant's contribution [88–90]. Gamification has been shown to contribute to motivation of participants [82] and it can be used to decrease the need for monetary incentive [91]. In this study, we consider how to design motivating functions which incorporate the concepts of interruption through notification and gamification.

### 3) Processing functions

In general, organizers intend to analyze or visualize urban environmental data collected with participatory sensing. Hence, participatory sensing platforms must support easy and quick access to this data. In the processing functions, the organizer defines methods of data processing to be used. The following functions are implemented in conventional platforms:

#### **Data editing (F6):**

Involves data cleansing and labeling.

#### **Data browsing (F7):**

Involves monitoring the status of data collection.

#### **Data export (F8):**

Involves exporting of collected data for more detail analysis or visualization using third-party tools.

Funf [78,92] and OpenDataKit [81] are designed as platforms that mainly focus on data cleansing and visualizing, as well as exporting. Additionally, these platforms support data processing in a variety of environments such as in the cloud

and on the smartphones used for data collection (endpoint devices). Thanks to the various processing functions in the platforms, many research projects in the world utilize them. Although most of them also support basic functions, there are differences in functions supported by other conventional platforms [76, 79, 80].

In this study, we implement all functions (F6–F8) as in Funf [78, 92] and OpenDataKit [81]. During implementation, we considered how to make the required technical skills for organizers and participants lower. Specific skills needed to operate existing platforms are described in the section below.

#### 4.2.2 Required Skills for Operation and Use

In participatory sensing, it is assumed that the organizer may be from a non-technical profession/background, e.g., they may be an administrative officer or urban planner, and ordinary people participate in the data collection. Thus, it is necessary to re-consider the skills required by the participatory sensing platforms for the organizers and participants.

Table 4.2 shows skill requirements for each conventional platform. Development skills (R1) and App distribution skills (R2) are required for organizers, where:

**Development skills (R1):**

Skills to develop the urban sensing system for the specific purpose.

**App distribution skills (R2):**

Skills to distribute client applications to participants' smartphone.

AWARE [76], Medusa [77], and OpenDataKit [81] have high extendability like a framework, but a high-level of programming skill is required for system construction (R1). In addition, most of the systems that require system development also require organizers to have the skills necessary to release applications on official stores such as Google Play and AppStore (R2). With web-based platforms such as MinaQn [79] on the other hand, deploying the applications is quite easy. However, continuously attracting users to the web application is required (R2).

App/Function management skills (R3) and Data processing skills (R4) are other skills that are required from participants, where:

Table 4.2. Overview of platform skill requirements.

Platforms	Skill requirements							
	Organizer skills				Participant skills			
	R1 Development	R2 App distribution	R3 App/Func. management	R4 Data processing				
AWARE [76]	As needed	-	Required	-	Required	-	-	-
Medusa [77]	Required	Required	-	-	-	-	-	-
Funf [78]	-	-	Required	Required	Required	Required	Required	Required
MinaQn [79]	-	Required	-	-	-	-	-	-
Ohmage [80]	-	-	Required	-	Required	Required	Required	Required
OpenDataKit [81]	As needed	Required	Required	Required	Required	Required	Required	Required
ParmoSense	-	-	-	-	-	-	-	-

**App/Func. management skills (R3):**

Managing applications and functions such as installation of applications and setting of functions to accomplish sensing tasks.

**Data processing skills (R4):**

Processing data collected using the participant’s device before uploading.

Since AWARE [76] and Ohmage [80], OpenDataKit [81] have many functions as a platform, the functions are divided into multiple applications and provided to participants. Also, Funf [78] requires participants themselves to set up the sensors (e.g., sensing interval) for each smartphone. Therefore, to construct the sensing environment that the organizer intended, knowledge and skills related to applications and functions are required for participants (R3).

Funf [78] and OpenDataKit [81] adopt a mechanism where they ask participants to perform data processing and cleansing. Such processing requires knowledge and skill to judge whether data is good or bad, so the burden on participants is substantial (R4).

Overall, conventional platforms require various skills for both organizers and participants to construct and operate the system. To realize participatory sensing systems that can be used by non-technical people, these problems must be addressed. In this study, we propose a new platform to resolve these problems.

## 4.3 Scenario-based Participatory Sensing Platform

We design and implement a novel participatory urban sensing platform, *ParmoSense*, to solve the problems mentioned in Section 4.2. *ParmoSense* aims to solve the following key challenges:

- C1: Limited support of essential functions needed for participatory sensing systems (listed in Table 4.1)
- C2: Difficulty of system construction and operation

*ParmoSense* allows organizers and participants to operate or contribute to participatory sensing systems without complex procedures or technical skills.

### 4.3.1 ParmoSense Basic Principles

ParmoSense is based on the following three basic principles.

#### **Principle #1. Modularized functions**

ParmoSense must achieve two contradictory requirements, easiness of system construction and diversity of available functions. Therefore, we employ the idea of modularizing functions inspired by existing research [76, 80]. ParmoSense provides the sensing, motivating and processing functions in Table 4.1 and 4.2, and these can be combined to form a participatory sensing system.

#### **Principle #2. Standardized participatory sensing system**

Conventional platforms require a high level of technical skills, e.g., programming skills for organizers to customize the platform for a specific sensing task. In order to deal with multiple purposes flexibly, ParmoSense is composed as a combination of modularized functions as well as the detailed settings of each function. We unify the combination and settings as a *scenario*.

A *Scenario* contains such information as the scenario name, description, sensing targets, sensing area, period, motivation method, etc. Through GUI-tool in ParmoSense, the organizer can generate the scenario easily. Based on the created scenario, ParmoSense automatically configures both a server system and a client application by distributing necessary information. Since scenarios can be created using any combination of functions and settings, logically, any kinds of participatory sensing system can be built with ParmoSense. Another important advantage of the scenario-based system is that users can participate in various participatory sensing projects through one client application, whereas in the past, each participatory sensing project required a dedicated application.

#### **Principle #3. Customizable motivation engine**

The most effective way of motivating participants depends on the purpose of sensing. To realize sustainable urban environmental sensing, ParmoSense has a *Motivation Engine* with a variety of motivation algorithms. The Motivation Engine provides the following functions for motivating participants:

**Motivation based on the behavior of an individual:**

- Granting incentives according to contribution
- Visualization of contribution

**Motivation for all participants, regardless of contribution:**

- Providing competition mechanisms such as rankings
- Sharing of experiences among participants

Additionally, by considering temporal/spatial information, e.g., the current time and the current position of the participant, it is possible to control the actuation timings of these functions. The optimal motivating algorithm according to the purpose of the organizer can be incorporated into the participatory sensing system by combining and customizing these functions.

### 4.3.2 ParmoSense System Overview

The architectural design of ParmoSense is shown in Figure 4.1. ParmoSense consists of three parts that have the following roles:

**ParmoSense Dashboard:**

A web application for organizers that can be used to create and distribute scenarios of participatory sensing systems, and to process collected data.

**ParmoSense Client:**

A client application for participants that can run various scenarios. By downloading and installing a scenario, it behaves as the corresponding sensing application.

**ParmoSense Server:**

A central system for integrated management of the scenarios created by organizers and automatically constructing a virtual server system for each scenario. It can collect sensing data from ParmoSense Client and generate feedback based on the analysis of the collected data.

Arrows in Figure 4.1 show contents transferred between organizer/participant and ParmoSense Server for each operation phase of participatory sensing. Each phase is defined as follows:

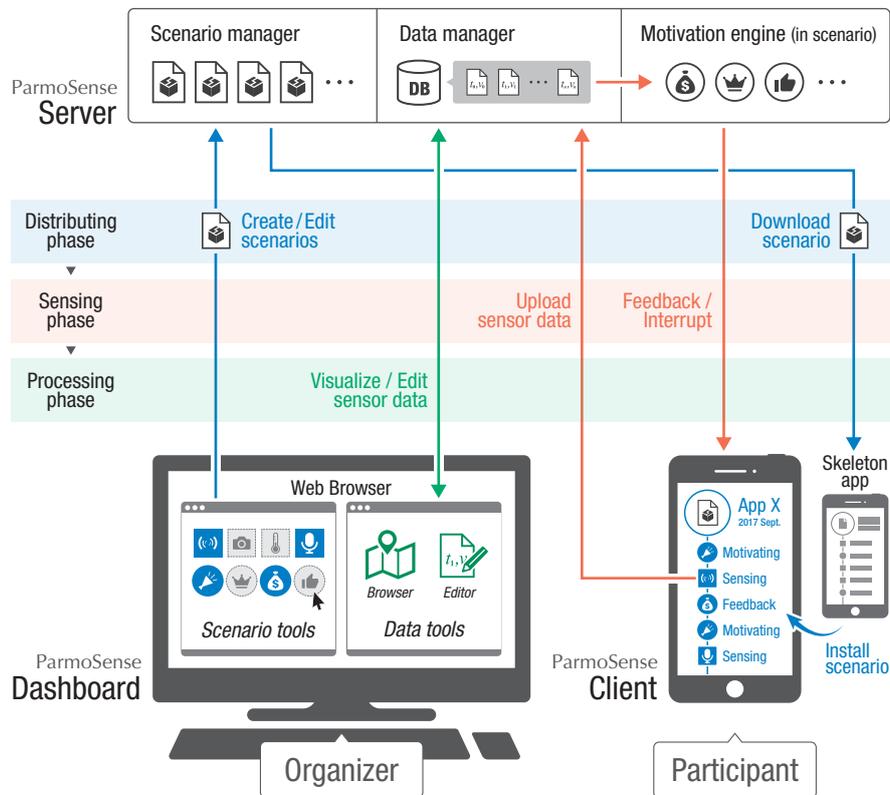


Figure 4.1. Design concept

1. **Distributing Phase:**

The phase for distributing the scenario created by the organizer to the participants via ParmoSense Server.

2. **Sensing Phase:**

The phase for giving feedback (e.g., monetary incentive) to the contribution of the participant such as uploading sensing data by participants.

3. **Processing Phase:**

The phase for editing, cleansing and visualizing the collected data.

Since ParmoSense is based on *Principles #1, #2* (Section 4.3.1), organizers can distribute sensing applications by simply exchanging scenarios with participants in Distributing Phase. It is therefore not necessary for each participant to manage

the applications by himself. Furthermore, thanks to *Principle #3*, in Sensing Phase, the feedback for motivating participants is created by the Motivation Engine using the collected data, and automatically provided to participants.

### 4.3.3 ParmoSense System Architecture

The concrete system configuration of ParmoSense is shown in Figure 4.2 (the overall system architecture) and Figure 4.3 (the system architecture for each scenario). In the following subsections, we will describe the ParmoSense Dashboard used by organizer, the ParmoSense Client used by participants, and the ParmoSense Server in more detail.

#### 1) ParmoSense Dashboard

An organizer carries out every operation, e.g., management of a participatory sensing scenario, processing of collected data on the ParmoSense Dashboard, a web application. It consists of *Scenario tools* (Figure 4.2 ①) and *Data tools* (Figure 4.2 ②).

Scenario tools, shown in Figure 4.4, provide many operations such as creating, editing and deleting participatory sensing scenarios, and browsing, activating and stopping scenarios. Figure 4.4 (a) shows the user interface for editing scenarios. The organizer can describe a scenario using the three kinds of functions, sensing, motivating and processing functions, mentioned in Section 4.2, without programming, through the GUI (*Principles #1, #3*). The scenario defined in Scenario tools is automatically converted to JSON format (see Listing C.1 in Appendix Appendix C.), and transferred between each part of ParmoSense.

When the scenario editing is completed, the virtual server system is automatically built depending on the scenario, and deployed by *Scenario Manager* (Figure 4.2 ③). At the same time, the QR code for downloading the scenario to ParmoSense Client is automatically generated. Figure 4.4 (b) shows the user interface for browsing/managing the scenario created by organizers. Scenarios currently in progress/stopped are indicated by blue/gray respectively, and these statuses and the scenario settings can be changed by using the GUI.

Data tools, shown in Figure 4.5, provide the functions for processing and visu-

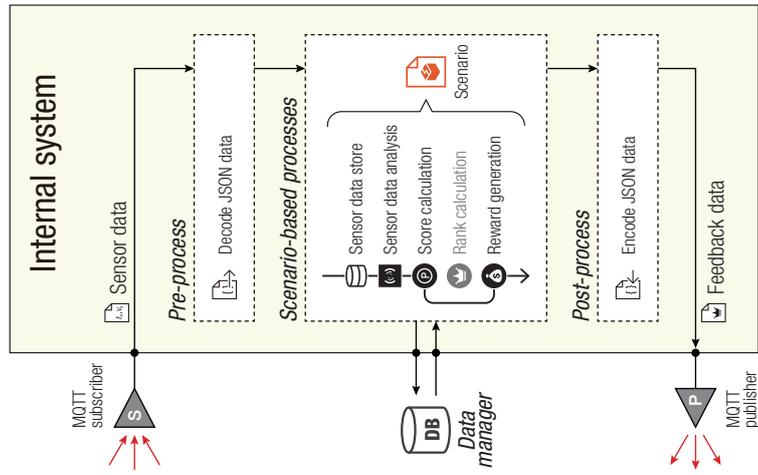


Figure 4.3. Internal system (Figure 4.2 (5))

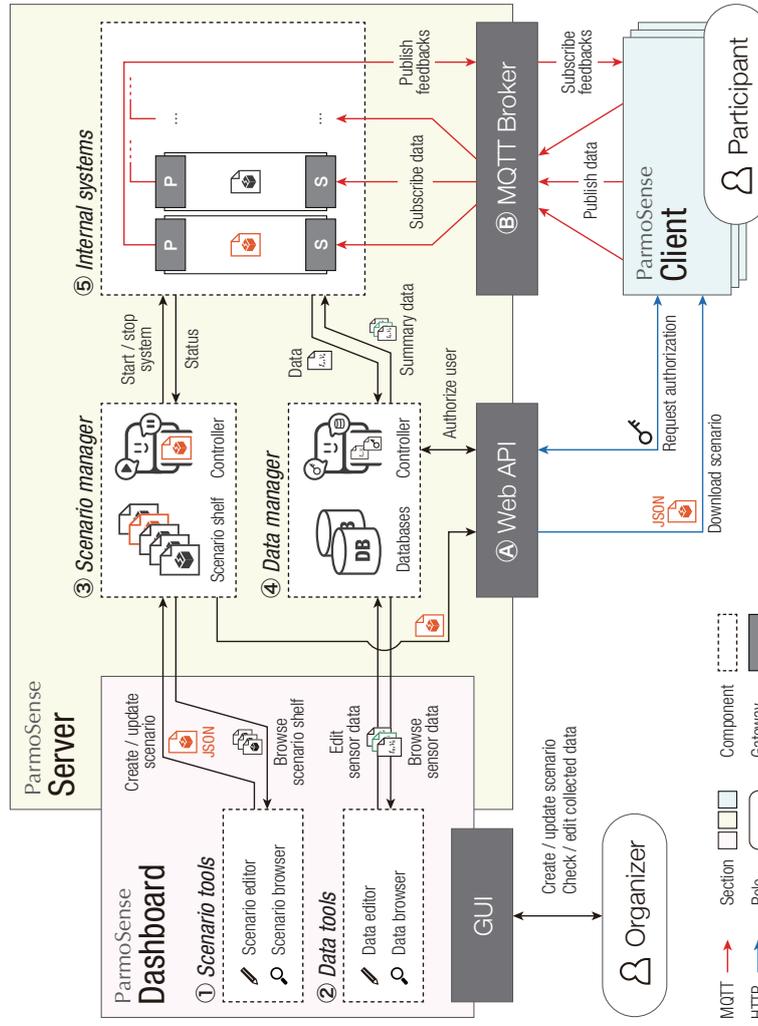
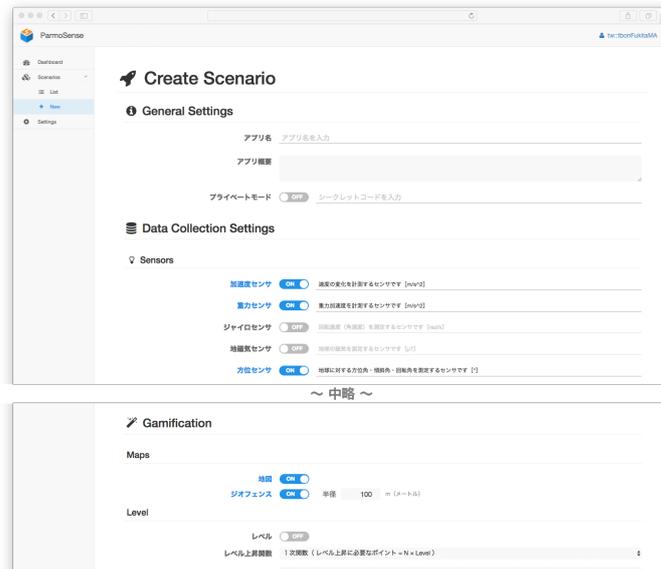
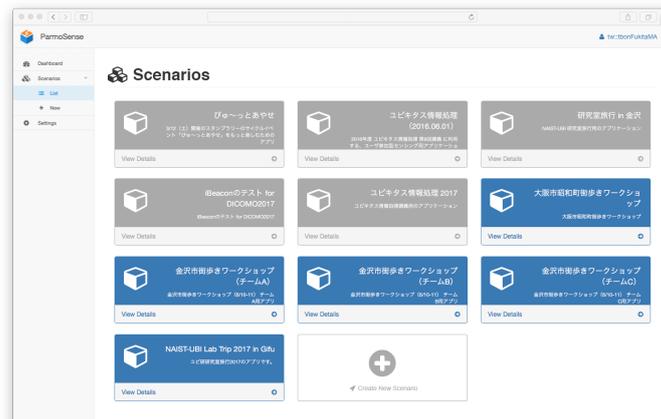


Figure 4.2. System architecture of ParmoSense



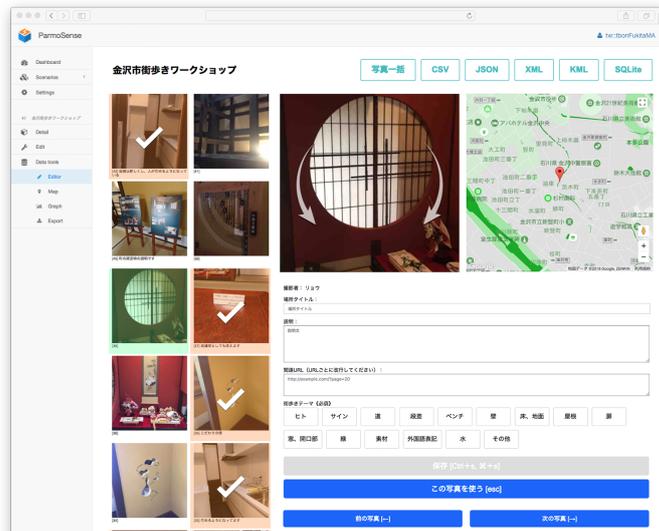
(a) Scenario editor



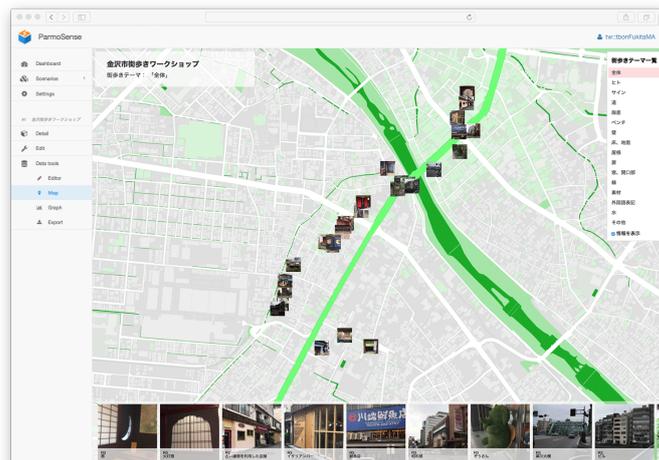
(b) Scenario browser

Figure 4.4. Scenario tools interface of ParmoSense Dashboard (Figure 4.2 ①)

alizing data aggregated into *Data Manager* (Figure 4.2 ④). The user interface for editing the collected data is shown in Figure 4.5 (a). The organizer can improve the quality of the data by editing/excluding inappropriate data from the collected data. The organizer can also add labels to the data. The processed data can be exported in the form of JSON, CSV, RDB etc. The user interface for visualizing the collected data is shown in Figure 4.5 (b). The organizer can



(a) Data editor



(b) Data browser (Map data: Google, Zenrin)

Figure 4.5. Data tools interface of ParmoSense Dashboard (Figure 4.2 ②)

check the data in two ways: overlaying them on a geographical map, or sorting them in a time-series order.

## 2) ParmoSense Client

The participant performs all sensing tasks of ParmoSense through the ParmoSense Client smartphone application. ParmoSense Client runs on smartphones with An-

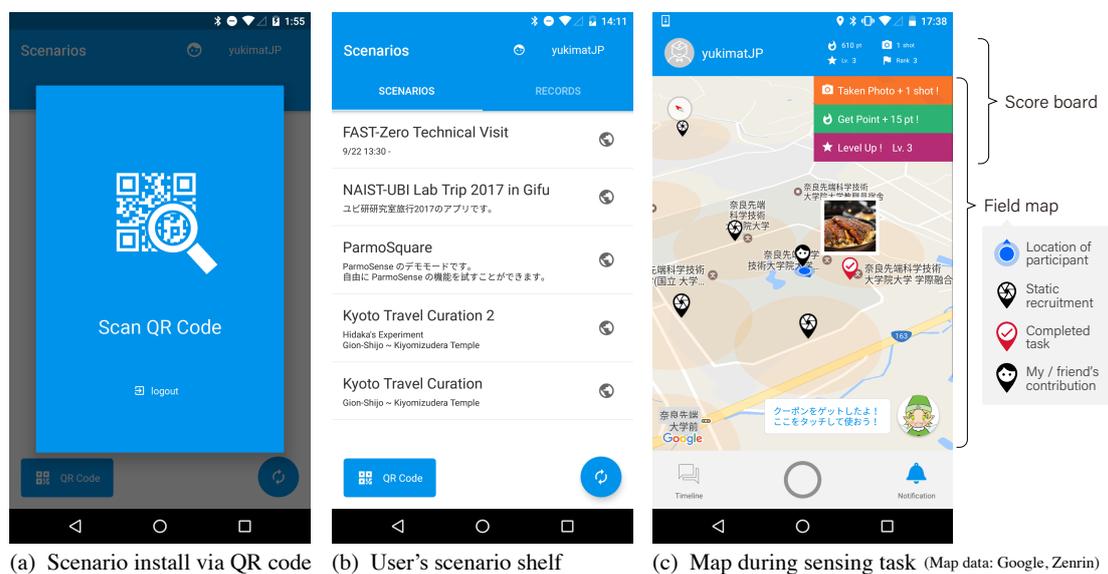


Figure 4.6. ParmoSense Client interface

droid OS or iOS, and it can be installed from general application stores (Google Play, AppStore). Since the behavior of participatory sensing system on ParmoSense is defined by a scenario (*Principle #2*), it can behave as various sensing applications by installing scenarios to ParmoSense Client.

Figure 4.6 (a)(b) shows the user interface for the scenario installation. The participant performs the following steps in order to install the application:

1. Log in on ParmoSense Client via Google Authentication.
2. Scan the scenario QR code<sup>\*1</sup> by a smartphone camera (Figure 4.6 (a)).
3. The participant confirms participation in sensing.

This procedure makes it easy to install scenarios. All these are done via the *Web API* shown in Figure 4.2 (A). Participants can participate in multiple scenarios. The scenarios that have been installed and performed in the past are listed, as shown in Figure 4.6 (b). This makes it easy for a participant to join the same scenario again.

<sup>\*1</sup> An organizer can get QR codes of scenarios from ParmoSense Dashboard, and print it out for participants.

Figure 4.6 (c) shows an example of the ParmoSense Client interface after installing the scenario and participating in sensing tasks. Requests of static sensing tasks are shown as pins on the map, and the participants can carry out the task at this place and acquire the incentive accordingly. The user score is displayed in the upper right area. It is a means of providing feedback to the participant through gamification and visualization of contributions. To provide the feedback and the execution status of tasks are reflected in real time according to the participant's and other participants' actions, ParmoSense have adopted MQ Telemetry Transport (MQTT) [93], which realizes many-to-many and real-time communication, as a communication protocol. The communication using MQTT is done via *MQTT Broker* shown in Figure 4.2 (B).

### 3) ParmoSense Server

ParmoSense Server consists of three parts: *Scenario Manager*, *Data Manager* and *Internal systems* shown in Figure 4.2 (3), (4) and (5) respectively.

Scenario Manager plays the role of storing the scenario created on ParmoSense Dashboard, and constructing and managing Internal system in accordance with the scenario. The Internal system is an instance executed on a server which communicates with a ParmoSense Client. By automatically constructing this Internal system for each scenario and forming Internal systems, various scenarios in ParmoSense can be created without programming. Scenario Manager monitors the operation status of the Internal system, and the organizer can start or stop the operation. It also detects unexpected troubles in the Internal system, and it stops or restarts them. The data collected by participants is aggregated in Data Manager, and this data is used for calculating the participants' score, visualizing on the map and so on.

Figure 4.3 shows the mechanism of the Internal system. Scenario Manager builds the Internal system by incorporating the module program of corresponding functions (Sensing Functions, Motivating Functions, Processing Functions) based on the scenario that an organizer created. Internal system communicates with ParmoSense Client via MQTT as described above. If a participant sends (publish) the sensing data, the corresponding Internal system receives (subscribe) the data collected by participant's smartphone sensors, and processes using modularized

functions (e.g., analysis of data, calculation of ranking) which are described in the scenario. According to these results, response data is generated and published to all participants who should be informed.

### **4.3.4 Functions Support**

ParmoSense comprehensively supports functions investigated in Table 4.1. In this section, we outline the support status of each function.

#### **Implicit-sensing (F1)**

ParmoSense supports data collection from sensors embedded in smartphones. The types of supported sensors are shown below:

- Position sensor (e.g., GPS)
- Environmental sensors (e.g., light sensor, barometer)
- Inertial sensor (e.g., accelerometer, gyroscope)
- External sensor device (heart-rate sensor)

It also supports data collection of Bluetooth Low Energy (BLE) scan logs of peripheral devices (e.g., iBeacon, other smartphones). For all these sensors, detailed configuration such as measurement interval and enabling/disabling of background measurement can be set on Scenario Editor of ParmoSense Dashboard by the organizer.

#### **Explicit-sensing (F2)**

ParmoSense supports various kinds of data collection methods that cannot be collected by sensors embedded in a smartphone. One of them is photo uploading. When taking photos and uploading them, participants can provide additional data such as explanatory texts of the photo taken, GPS position, and other data obtained by Implicit-sensing. Questionnaires are also provided. Different question types such as binary questions (YES/NO question), multiple-choice questions (up to four options), and questions that require photo uploads and explanatory text, are supported. These questions can be chained together to support a step-by-step questionnaire.

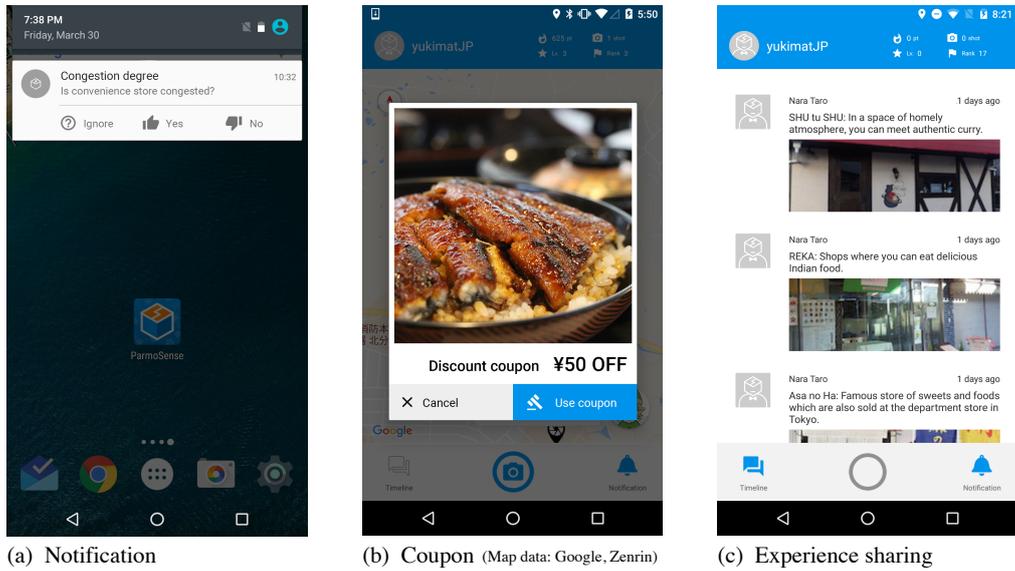


Figure 4.7. Examples of motivating functions

### Static/Dynamic request (F3)

ParmoSense supports both static and dynamic requests for soliciting contributions on sensing tasks. In participatory sensing for urban environments, there are many requests based on geographical information. Static requests place each task as a checkpoint on a map as shown in Figure 4.6 (c), which allows participants to easily find the tasks to be performed. For dynamic requests, we support informing participants of task requests through notifications as shown in Figure 4.7 (a). Organizers can generate and request tasks from specific participants, for example, by setting that any person who is detected entering a certain area is notified. Location information can be obtained using region monitoring technology such as Geofence and iBeacon).

### Monetary/Non-monetary incentives (F4)

ParmoSense supports both monetary and non-monetary incentives to compensate participants for their contributions. The monetary incentives supported are discount coupons that can be used at restaurants, cafe and so on, as shown in Figure 4.7 (b). Since the discount coupon is linked with purchasing behav-

ior, it is effective for motivating participation in scenarios such as sightseeing. For non-monetary incentives, ParmoSense supports the gamification mechanism. This includes awarding points for a contribution, competition mechanisms such as comparing a participant's degree of contribution to that of other participants, and virtual level-up elements by repeating the contribution.

### **Feedbacks by visualization (F5)**

ParmoSense also supports feedback not included in game features, for example, visualization of own contributions, and data/experience sharing. Visualization of own contributions is of two types: (i) plotting the pins of contributions on the map and (ii) scoring contributions with the non-monetary incentives detailed above. In addition, a data sharing function to share/visualize data such as participants' experiences and what different participants viewed, heard, or sensed (environmental conditions) is provided. For example, ParmoSense can plot sensing data uploaded by other participants on the map and share them on a timeline as shown in Figure 4.7 (c).

### **Editing of collected data (F6)**

ParmoSense helps organizers to pre-process collected data before detailed analysis. For example, it provides functions such as cleansing unnecessary data, selecting data to be used, and labeling the data. Also, since these data processing functions are designed to protect the original data, it can be restored at any time.

### **Browsing of collected data (F7)**

ParmoSense provides visualization tools for instant and easy checking collected data. There are many types of tools such as a tool for plotting data collected by all or each participant on the map, and a tool for displaying all data as a list. These tools can be used at any time even while sensing tasks are in progress.

### **Export of collected data (F8)**

ParmoSense supports various data output formats. Data analysis data can be output in CSV, JSON, XML, RDB (SQLite), etc., so that analysis can be started

immediately. Moreover, when exporting to a third-party visualization tools (e.g., Open Street Map [94], Google Earth [95], Cesium [96]), it is possible to output in KML<sup>\*2</sup> or GPX<sup>\*3</sup>.

## 4.4 Evaluation

ParmoSense has tackled two challenges to be solved: “limited support of essential functions in existing participatory sensing systems (C1),” and “difficulty of system construction and operation (C2).” In this section, we use the *number of functions provided* and *ease of system construction and operation* as the metrics, and we evaluate the performance of ParmoSense compared with conventional platforms.

### 4.4.1 Workload for System Operation

We evaluated the difference in the development cost and the operation cost of ParmoSense compared to conventional platforms [76–81]. We used the time cost (Preparation Time) for starting operation of the system as the metric of comparison. The definition of Preparation Time is as follows:

Preparation Time ( $T$ )

Preparation Time ( $T$ ) is the total time required from the start of developing the participatory sensing system to the start of sensing tasks, which is calculated by the sum of sub tasks (Equation (4.1)).

$$T = \sum_{i=1}^5 t_i \quad (0 \leq T) \quad (4.1)$$

Where  $t_1 \sim t_5$  is the relative estimated time required for each task, as shown in Table 4.3. As reference, we defined the time required to install one application from a general application store, e.g., Google Play or AppStore as  $t = 1$ .

Estimated Preparation Time on each platform is shown in Table 4.4 and the x-axis of Figure 4.8. The filled circle (●) represents the case where the simplest

<sup>\*2</sup> <https://developers.google.com/kml/>

<sup>\*3</sup> <http://www.topografix.com/gpx.asp>

system on each platform was constructed. AWARE [76], OpenDataKit [81] and Medusa [77] can be extended by programming as necessary. However,  $t_1$  will increase linearly with the development of functionality.

Therefore, ParmoSense belongs to the group which needs lowest Preparation Time, and can be operated as easily as other platforms in the same group. Also, conventional platforms [76,77,81] suffer from the problem where the time required for expansion tends to be long when trying to extend the systems' functionalities. In contrast, ParmoSense is designed to cover the necessary functions in advance, and thus, although preparation time required is equivalent to other platforms, it achieves higher functionality.

#### 4.4.2 Variety of Functions

We evaluated the diversity of functions that ParmoSense provides, compared to conventional platforms [76–81]. As a metric for comparison, we use the following fulfillment status of functions (Function Score):

Function Score ( $S$ )

Table 4.1 shows all the functions supported by each of the existing participatory sensing platforms. The score of each function  $s_1 \sim s_8$  is determined by the implementation status of each function (F1 ~ F8) of Table 4.1, and Function Score ( $S$ ) is calculated by the sum of them (Equation (4.2)).

$$S = \sum_{i=1}^8 s_i \quad (0 \leq S \leq 8) \quad (4.2)$$

$$s_i = \begin{cases} 1 & \text{if } F[i] = \checkmark \\ 0.5 & \text{if } F[i] = \triangle \\ 0 & \text{otherwise} \end{cases}$$

The y-axis of Figure 4.8 shows the estimated Function Score. While conventional platforms have moderate scores ( $S = 4 \pm 1$ ), ParmoSense has a higher score ( $S = 8$ ) due to comprehensively supporting all the functions provided in existing platforms, and incorporating motivating functions, which are missing on all existing platforms. Furthermore, the advantage of ParmoSense will be even higher if we consider the effect brought by the combination of these functions. For

Table 4.3. Preparation Time for Each Task

	Contents of work	Preparation time [h]	Role
$t_1$	development (programming)	0, 8 <sup>*a</sup>	organizers
$t_2$	development (GUI editing)	0, 4 <sup>*e</sup>	
$t_3$	publishing to app store	0, 16 <sup>*b</sup>	
$t_4$	installing application	0, 1 <sup>*c</sup> , 2 <sup>*d</sup>	participants
$t_5$	configuring functions	0, 1 <sup>*e</sup>	

<sup>\*a</sup> calculated using LOC of source code.

<sup>\*b</sup> calculated based on time for becoming available on the general store.

<sup>\*c</sup> installation of single application (reference time). <sup>\*d</sup> installation of multiple application.

<sup>\*e</sup> calculated by comparing with  $t_1$ ,  $t_3$  and  $t_4$  relatively.

Table 4.4. Breakdown of Preparation Time

Platforms	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$	$T$
AWARE [76]	- <sup>*a</sup>	4	-	2	1	7
Medusa [77]	8 <sup>*a</sup>	-	16	1	-	25
Funf [78]	-	-	-	1	1	2
MinaQn [79]	-	4	-	-	-	4
Ohmage [80]	-	-	-	2	1	3
OpenDataKit [81]	- <sup>*a</sup>	4	16	2	1	23
ParmoSense	-	4	-	1	-	5

<sup>\*a</sup> Additional time is required for developing new functionalities.

example, ParmoSense can provide combinations of multiple motivating functions, such as gamification, non-monetary incentives, and interruptions, which can not be provided on other existing platforms, and which may be more effective than using a single motivating strategy/function, since they cater for more participant preferences.

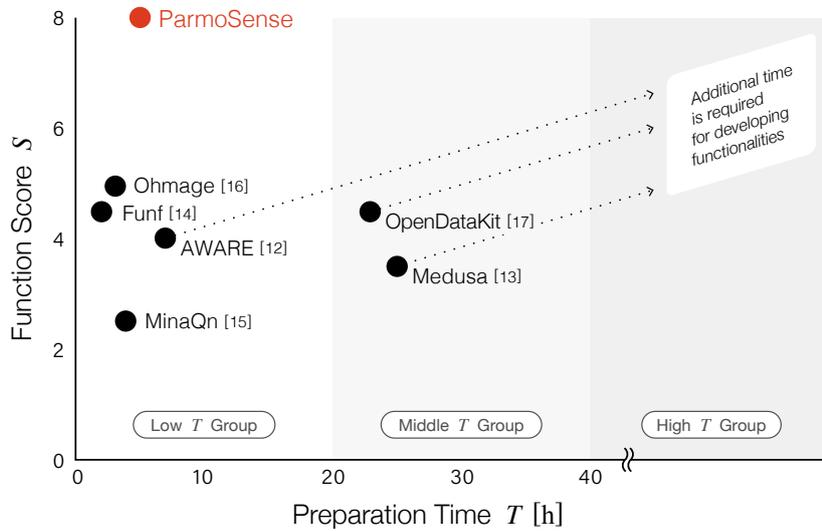


Figure 4.8. Comparison graph

## 4.5 Case Studies

In Section 4.4, we evaluated the number of functions provided, and the time cost of using ParmoSense, compared to conventional platforms. Prior to the evaluation, we released ParmoSense to general application stores (Google Play, AppStore). In two years after, we conducted 17 case studies. We collaborated with various organizers to design and build 17 scenarios, and then deployed them via the client application. Members of the general public who had downloaded the application acted as participants.

Our aim in doing these case studies was two-fold. First, we wanted to validate that our implementations of the sensing, processing and motivating functions on ParmoSense would work well in real-world participatory sensing tasks. Second, we wanted to determine how effective the functions were in motivating participation and in supporting organizers to collect required data and to extract the required information. Such knowledge would allow us to improve the functions, or to recommend additional functions to be included in participatory sensing systems. Table 4.5 shows the details of each scenario (start date, period, number of participants, and embedded functions). In following subsections, we discuss the sufficiency of ParmoSense functions in each scenario.

Table 4.5. Overview of case studies.

Scenario No.	Functions											Start date	Period	Number of participants	
	sensing functions		motivating functions			processing functions									
	F1 Implicit	F2 Explicit	F3 Request	F4 Reward	F5 Feedback	F6 Editing	F7 Browsing	F8 Export							
S1	✓	✓		✓	✓		✓		✓				Feb. 9th, 2016	0.5 hours	15
S2	✓	✓			✓		✓		✓				Feb. 17th, 2018	1 day	9
S3	✓	✓		✓	✓		✓		✓				June 1st, 2016	0.5 hours	17
S4	✓	✓			✓		✓		✓				Dec. 17th, 2016	2 days	14
S5	✓	✓			✓		✓		✓				June 8th, 2017	1 day	6
S6	✓	✓			✓		✓		✓				June 10th, 2017	1 day	20
S7	✓	✓		✓	✓		✓		✓				June 5th, 2016	2 days	19
S8	✓	✓			✓		✓		✓				Nov. 22nd, 2016	3 days	14
S9	✓	✓		✓	✓		✓		✓				Nov. 26th, 2017	1 day	30
S10	✓	✓		✓	✓		✓						Jan. 21st, 2016	10 days	12
S11	✓	✓					✓		✓				Apr. 20th, 2017	14 days	83
S12	✓	✓		✓	✓		✓		✓				Mar. 12th, 2016	1 day	10
S13	✓	✓		✓	✓		✓		✓				Nov. 13th, 2016	1 day	48
S14	✓	✓		✓	✓		✓		✓				Feb. 25th, 2017	1 day	25
S15	✓												June 11th, 2017	2 days	18
S16	✓												July 24th, 2017	4 hours	100
S17	✓												Sept. 24th, 2017	6 hours	200

### 4.5.1 Overview of Case Studies

We categorize case studies into four groups according to the type of sensing tasks involved and provide an overview of each group in this section. We then describe how well the ParmoSense functions performed for each scenario.

#### Urban-data collection during workshops (S1–S6)

S1–S6 are scenarios for workshop-style events, such as a Mapping-party [97, 98], which is widely used in organizations such as OpenStreetMap [99]. The overview of each scenario is as follows:

##### Scenario S1–S2 (Mapping-party)

These scenarios were aimed at collecting unmapped geographical data. In this event, we collected information on trees (names and positions) in our university campus.

##### Scenario S3 (FixMyStreet)

This scenario was for collecting dynamic geographical data such as road breakages, graffiti, and street lamp failures by imitating the mechanism on FixMyStreet [100].

##### Scenario S4–S6 (Urban-planning)

These scenarios involved surveying existing point of interests (POIs) such as local buildings, public facilities, and nature, for urban planning.

Through deploying these scenarios, we obtained the following knowledge:

- Because participants were willing to attend the event by themselves, we confirmed that the set of motivating functions on ParmoSense are enough for getting sufficient contributions from the general public.
- The following functions of processing functions worked effectively:
  - Data labeling function
  - Cleansing function for unnecessary data
  - Automatic mosaic function by face recognition
  - Data export function

## **Urban-data collection during sightseeing (S7–S9)**

S7–S9 are scenarios for sightseeing participatory sensing, e.g., with an information sharing tool for tourists, or an urban sensing system mimicking a tourist guide. The overview of each scenario is as follows:

### **Scenario S7 (Experience-sharing)**

This scenario involved sharing discovered sightseeing spots among a group of tourists. To encourage positive posting among participants, we used points as the motivating function (F4).

### **Scenario S8 (Tourist-guidance)**

This scenario involved collecting data such as photos and behavior logs from tourists, while providing sightseeing information through a virtual tour guide. We provided a data editing function (F6) for organizers to edit collected data after the event.

### **Scenario S9 (Multi-type-requests)**

The scenario was for comprehensively collecting environmental data of sightseeing spots, by requesting it in various ways from tourists. To encourage active contribution of participants, we provided all the available motivating functions such as static/dynamic request (F3), points and coupons (F4).

Through these case studies, we observed the following on the effectiveness of the implemented functions for scenarios belonging to this category:

- In the case of sightseeing, motivating functions were essential because tourists participated in the scenario opportunistically.
- Collecting factors, e.g., points and sharing information with each other, were more effective than the competitive factors, e.g., score and ranking, at motivating participant’s continuous participation because “sightseeing” was their primary purpose and collecting factors were more relevant.

## **Urban-data collection in daily life (S10–S11)**

The data suitable for collecting by participatory sensing are often “continuous” and “long-term” data, because participatory sensing is a sustainable sensing

mechanism to realize comprehensive spatio-temporal data collection without any infrastructure due to the use of the many mobile devices dispersed in the city. S10 and S11 were scenarios for collecting such long-term and continuous data. The overview of each scenario is as follows:

### **Scenario S10 (Static-request)**

This scenario was for periodically collecting information that changes day by day, e.g., bulletin board information at a facility, and temperature of a place. To encourage participation, we set a limit on the number of participant contributions resulting from static requests (F3) that would be accepted, which is similar to the Auction mechanism [77].

### **Scenario S11 (Dynamic-request)**

This scenario was for conducting questionnaires linked to location information by interrupting participants. Push notifications were used, and participants' location and behavior were taken into consideration, to provide dynamic requests (F3). No maps, incentives, or visualizations were included.

Through the case studies involving these scenarios, we concluded the following regarding the effectiveness of ParmoSense functions intended for urban-data collection in daily life:

- In the case of S10, due to the use of static requests through placing pins on a map, the achievement of sensing depended on the active and continuous contribution of participants themselves. The following functions worked efficiently when motivating contribution:
  - Virtual non-monetary incentives (e.g., point, ranking, level)
  - Monetary incentives
  - Visualization of contributions of the participant and of other participants
- Restricting the number of contributions on static requests was effective for encouraging *continuous* participation because it prevented point inflation, but it also caused the leaving of some participants.

- In the case of S11, we found that contribution of participants improved when the timing of requests was based on participant’s location and behavior, e.g., participants were more likely to respond to requests if the location of the sensing task was near to their present location.

### **Human-behavior investigation on events (S12–S17)**

ParmoSense can not only collect urban environmental data, but also data of “people” existing in the city. S12–S17 were scenarios created for investigating human behavior during various situations: events, daily life, sightseeing and so on. The overview of each scenario is as follows:

#### **Scenario S12–S14 (Stamp-rally)**

These scenarios was for investigating the behavior of people who participate in the electronic “stamp-rally<sup>\*4</sup>”. A physical (electronic) stamp is put in a predetermined place, and when a visitor arrives, a stamp is stuck on a sheet (application). Reward is given based on the number of stamps accumulated. The non-monetary incentive include coupons, prizes, points and ranking as monetary/non-monetary incentives (F4). We provided a data browsing function for visualizing participants’ behavior (F7). Figure 4.9 shows examples of visualization of participants’ behavior.

#### **Scenario S15 (Sightseeing)**

This scenario was for collecting participants’ movement data linked with the location information, to investigate the participants’ behavior tendencies during sightseeing. To continuously sense tourists’ behavior, we used an implicit sensing function that ran in the background (F1).

#### **Scenario S16–S17 (w/external-sensor)**

These scenarios was used to collect participants’ movement and heart rate data linked to location information, to be used to construct a database for human behavior analysis. We added a function to connect a wearable sensor to external sensors (F1) via BLE, for collecting data.

---

<sup>\*4</sup> A stamp-rally is a motivating approach used to entice users to go around to certain places at events etc. [101].

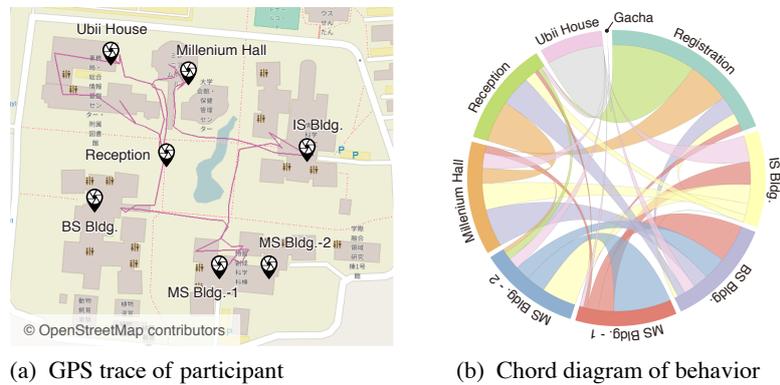


Figure 4.9. Result of human-behavior investigation

Through these case studies, we learned the following about participatory sensing for human-behavior investigation:

- To analyze the behavior of a person, there is a need to visualize behavior logs in various ways. We found that the functions provided by ParmoSense such as GPS trace, and chord diagram listings worked effectively.
- Since these scenarios were aimed at sensing of human behavior affected by the specific function, it was necessary to suppress interference of elements, which is except for the function to be evaluated, to people. We confirmed that ParmoSense can minimize the number of unnecessary functions because of module-based design.
- It is sometimes necessary to collect data with high frequency over a long term<sup>\*5</sup>. In this case, we found that functions such as continuous data acquisition in the background worked efficiently.

#### 4.5.2 Survey and Discussion

In the following sections, we summarize and discuss the results of the subjective surveys held in each case study. We asked each participant and organizer questions about the usability and performance of ParmoSense.

<sup>\*5</sup> S16 and S17 were scenarios that required acquiring the 9-axis sensor data at 100 Hz over several hours.

## 1) Overview

Participants in nine scenarios (S2, S4–S6, S9, S11–S14) were given the questionnaire. These scenarios had many ordinary people participating, and include at least one motivating function. The total number of participants who answered the questionnaire was 201. About 70 % of participants were in their 20s, however, various age groups ranging 10s to 80s were targeted. Attributes of participants are listed below:

- Citizens living in the local area, men and women of all ages (S2, S4, S13–S14)
- Undergraduate students majoring in architectural engineering in their 20s (unfamiliar with the information science) (S5, S6)
- Graduate students majoring in information science (S9, S11)

We also distributed questionnaires to the organizers of five scenarios (S2, S4–S6, S9). The skills of the organizers varied from those with high ICT skills like ICT engineer to those with low ICT skills like students majoring in fields other than IT. The attributes of the organizers are listed below:

- Employee of regional public facility (S2)
- ICT engineer (S4)
- Undergraduate students majoring in architectural engineering (S5, S6)
- Graduate students majoring in information science (S9)

## 2) Survey of participants

To evaluate the usability of ParmoSense Client, we asked participants the questions which are listed in Table 4.6. we asked **Q1** (“*Was ParmoSense Client easy to use?*”) with a 5-Point Likert scale (5: very easy to use ~ 1: very hard to use), and then asked for the reason in **Q2**. The total number of answers was 196. The breakdown of answers is shown in Table 4.6. The average score was 3.4 and about 40 % of participants answered “*easy to use.*”

Participants who answered “*easy to use*” had the following to say:

Table 4.6. Questionnaire items for participants and non-participants (S2, S4-S6, S9, S11-S14)

Item No.	Questionnaire Detail	Answer					Average
		1 (disagree)	2	3	4	5 (agree)	
<b>Q1</b>	Was ParmoSense Client easy to use? (S2, S4-S6, S9, S11-S14)	8 (5.0 %)	19 (11.9 %)	73 (45.6 %)	35 (21.9 %)	22 (15.6 %)	3.4
<b>Q2</b>	Why did you think so about the answer to <b>Q1</b> ? (S2, S4-S6, S9, S11-S14)	(Open-ended question)					-
<b>Q3</b>	Was the event using ParmoSense fun? (except S11)	- (0.0 %)	4 (3.9 %)	14 (13.6 %)	42 (40.8 %)	43 (41.7 %)	4.2
<b>Q4</b>	Why did you think so about the answer to <b>Q3</b> ? (except S11)	(Open-ended question)					-
<b>Q5</b>	What factors will encourage you to participate in experiments? (Non-participants of S11)	(Open-ended question)					-

- “We could operate intuitively because of simple application design.” (S4, S6, S14)
- “Map visualization function is helpful for understanding the collected data having position information.” (S4, S5, S6)

Those who answered “hard to use” gave the following comments:

- “It was difficult to use ParmoSense application because I was not used to smartphones.” (S4, S13)
- “ParmoSense application was not suitable for long-term use because this application consumes battery more than expected.” (S4, S5, S6)
- “Unused functions should be invisible.” (S11)

Overall, we found that using ParmoSense was easy for participants who use smartphones on a daily basis regardless of age. In the case of local events where elderly persons also attended, some seniors felt that it was difficult to use a smartphone. However, this is not a particular problem of ParmoSense. As the use case diversifies, the information displayed on the screen also diversifies. Thus, to solve this issue without programming skills, it is necessary to modularize the screen configuration of the application and also to enable screen layout design in Scenario tools according to the use case.

Next, we asked **Q3** (“Was the event using ParmoSense fun?”) with a 5-Point Likert scale (5: very fun  $\sim$  1: not fun at all) in target scenarios except for S11. Additionally, we also asked **Q4** to solicit free responses from participants on how motivating functions such as Reward and Visualization affected their sensing behavior. The total number of answers was 103. The breakdown of answers is shown in Table 4.6. The average score from the responses was 4.2 and about 80 % of participants answered “fun.” In **Q4**, we obtained the following comments:

- “It is pleasant to see the behavior of other people and other groups by checking the pins on the map and timeline in real time.” (S4, S10)
- “Ranking function and leveling function makes it fun and encouraging.” (S4, S5, S9)
- “This application increased the fun of the stamp-rally.” (S13, S14)

### 3) Survey of non-participants

We also provided questionnaires to non-participants in the free participation scenario (S11) in order to explore the reasons why they did not participate in the sensing tasks.

In scenario S11, 35 of 118 candidates did not respond to the request to attend our experiment. We asked these 35 non-participants the open-ended question **Q5** (“*What factors will encourage you to participate in experiments?*”). The factors that promoted participation were related to ease of participation (17 people), battery consumption concerns (10 people), incentives (13 people), and benefit or convenience (12 people) respectively. Furthermore, four people were concerned about privacy, such as “*requiring personal information (e.g., e-mail address)*” and “*cannot be anonymous.*”

Almost half of the non-participants pointed out that the number of steps they needed to take in order to participate was inconvenient. To minimize the steps to participate in multiple scenarios, we first required participants to install Par-moSense Client. After installing the application, participants read a QR code of each scenario in order to participate in that particular scenario. However, people who participated in only one scenario felt that this is complicated and that having a single, pre-configured application for a specific scenario might be easier.

Also, to make the registration and login process easier, we adopted Google Authentication (using Gmail address) and supported auto-login. This is a standard method used in many applications. However, some participants felt that our application could collect private data such as the e-mail address. In the future, we aim to address this by introducing an anonymous participation mode that does not require participant registration.

### 4) Survey of organizers

To evaluate the effect of Par-moSense from the view of organizers and the usability of processing functions, we conducted a questionnaire with organizers. The questions for organizers are listed in Table 4.7. **Q6–Q11** were answered with 4-Point Likert scale (4: strongly agree ~ 1: strongly disagree), **Q12** was a free-response question.

Table 4.7. Questionnaire items for organizers (S2, S4-S6, S9)

Item No.	Questionnaire Detail	Answer				Average
		1 (disagree)	2	3	4 (agree)	
<b>Q6</b>	Was ParmoSense easy to introduce to the event?	-	-	3	2	3.4
<b>Q7</b>	Could you collect the desired data?	-	-	1	4	3.8
<b>Q8</b>	Were participants' motivation for attending to events improved by using ParmoSense? (e.g., ranking, coupons)	-	-	3	2	3.4
<b>Q9</b>	Was Data tools easy to use?	-	1	2	2	3.2
<b>Q10</b>	Were the data outputted by the data output function easy to use secondary diversion and data processing?	-	1	3	1	3.0
<b>Q11</b>	Do you want to use ParmoSense again in future similar events?	-	-	2	3	3.6
<b>Q12</b>	Why did you think so about the answer to <b>Q11</b> ?	(Open-ended question)				-

To evaluate the organizer’s burden indicated in Figure 4.1 – Distributing-phase, we asked about **Q6** (Ease of creating and distributing scenarios) to organizers. Two people answered “4” and the other three answered “3,” giving an average score of 3.4. Although the ICT skills of the organizers were quite different, all of them felt that ParmoSense was easy to use for participatory sensing.

Next, we asked **Q7** (Whether they were able to collect desired data) and **Q8** (Performance of motivating function), to get feedback from organizers about Figure 4.1 – Sensing-phase. Four organizers answered “4” and one answered “3” in **Q7**, while two answered “4” and the other three answered “3” in **Q8**. Consequently, we can say that ParmoSense allows organizers to collect desired data and to motivate participants to contribute the sensing continuously, from a organizer’s objective perspective.

To evaluate Figure 4.1 – Processing-phase, we asked the organizers about the **Q9** (Usability of Data tools) and **Q10** (Availability of output data). Two organizers answered “4” and two answered “3” and another organizer answered “2” for **Q9**, while one organizer answered “4” and three answered “3” and other one answered “2” for **Q10**. The average scores were 3.2 and 3.0 for **Q9** and **Q10**, respectively. According to these results, we can confirm that the Data tools of ParmoSense work substantially fine.

Additionally, we interviewed the organizers who answered “2” in **Q9** and **Q10**, respectively. The organizer whose answer for **Q9** is “2” said “*Unexpected data was downloaded when exporting data for each tag after tagging data.*” Therefore, as an additional question, we asked “*Is it easy to use if data could be successfully downloaded?*,” and we got the answer “*I think so.*” This shows that although it is necessary to improve the flexibility in the export function, it seems that the usability of Web editor meets specific service standards. Next, the organizer whose answer for **Q10** is “2” said “*When doing Web visualization, the operation became heavy and the PC froze many times.*” About 700 photos were collected in this organizer’s event. This data amount was too large to display all photos at once. In the future, it is necessary to set an upper limit on the number of photos that can be displayed, or reduce the image size according to the number of images before uploading.

Finally, three organizers answered “4”, two answered “3” in **Q11** (Whether to

use ParmoSense again). When we asked the reason in **Q12**, the following answers were obtained:

- “It can find places where the participants are interested in.”
- “It can easily collect data with location information and can edit detailed information later. In addition, by visualizing the data, participants can understand how the data is used intuitively.”

From these results, we found that ParmoSense is useful for human-behavior analysis and feedback of results to participants and it is easy to edit data.

## 4.6 Conclusion

Due to mechanism of the *Human-in-the-loop sensing* framework established on *cooperation* in which depends on the voluntarism of general people, the sustainability of this framework is a critical challenge in real-world operation. In this study, we have focused on the mutual linkage with the people in civic tech community, and designed and implemented a participatory sensing platform, named ParmoSense, that allows them to construct, customize, operate *Human-in-the-loop sensing* systems easily and quickly regardless of their ICT skills. To achieve this, we employ two features: *modularization of functions* and *scenario-based participatory sensing system description*. To evaluate ParmoSense, we conducted comparison tests with conventional platforms [76–81]. First, we confirmed that ParmoSense provides higher variety of functions than the conventional platforms, which solves challenge C1. We also found that the time required for system preparation using ParmoSense can be suppressed to a great extent compared with those platforms, which solves challenge C2. From both perspectives therefore, ParmoSense shows the best cost-performance. In addition, through 17 case studies over two years, we confirmed that ParmoSense could be compatible flexibly with various sensing tasks, motivate methods, and data processing. Moreover, ParmoSense can suppress the burden of organizers and participants in system operation, and we found that it has higher cost performance than any conventional platforms. On the other hand, we also found ParmoSense has not been a “magic bullet” solution that can be used in every situation. ParmoSense should

improve privacy and security by allowing organizers to control them based on the sensing purpose. In addition, to operate scenarios that involve a more significant number of participants, robustness and scalability should be guaranteed based on distributed processing and prediction of score update and so on.

# 5 Conclusion

## 5.1 Summary

In this dissertation, we presented the new urban environment sensing framework, named *human-in-the-loop sensing*, based on involving human beings to the sensing system. The motivation for our work is the demand for next generation context-aware system, that comprehensively and dynamically understanding contexts dispersed in the urban space. As the contexts which should be recognized, we emphasize the importance of subjective contexts, e.g., emotional status and satisfaction level, in addition to conventional objective contexts, e.g., crowdedness, network connectivity, and temperature. Hence, we have organized current urban sensing system methodologies as the human involvement model, and placed a *human-in-the-loop sensing* framework which consists of *human as a device carrier* and *human as a sensor* as the scope of this study. To realize a *human-in-the-loop sensing* framework, two challenges have been tackled in this dissertation: 1) How to extract objective and subjective contexts?; 2) How to sustainably operate *human-in-the-loop sensing*?

For the first challenge, we presented two urban environment sensing projects with different use cases: NightRoadScanner and EmoTour. Each system is based on *human as a device carrier* and *human as a sensor* concept respectively. The contribution of each system is shown as follows.

- The first system, NightRoadScanner described in Chapter 2 has provided the objective context estimation method based on *human as a device carrier* concept. We work on a safety level assessment system for sidewalks at night based on measurement of the illuminance of streetlamps by using a light sensor embedded a smartphone owned by general people. To estimate accurately using smartphone embedded sensors which have various sensor

characteristics, we have proposed a method to relatively correct data measured by unknown devices using the data measured by devices whose sensor characteristics is known. Through experiments in the real-world condition, we confirmed the proposed method reduced the estimation error for 90 % samples of streetlamps. Regarding the estimation accuracy, with our proposed method, the absolute error has been improved around 75 % compared with using raw data, and even improved around 20 % compared with applying our previous method. Finally, we have found our method can estimate illuminance of streetlamp with 0.54 lx averaged absolute error.

- The second system, EmoTour described in Chapter 3 has provided the subjective context estimation method based on *human as a sensor* concept. We have proposed a new model for quantitatively estimating both emotion and satisfaction of tourists by employing multiple modalities obtained from the unconscious and natural user actions. To estimate the satisfaction level, a user review is often used, however, it has the potential risk of biased ratings due to the imbalanced level of motivation. Also, some emotional status estimation systems have been proposed, but, they have not been evaluated at actual sightseeing situation. We employ the combination of behavioral cues and audiovisual data collected by an eye gaze tracker, physical activity sensors and smartphone. Then, we evaluated our model through experiments with 22 participants in a tourist domain (i.e., in a real-world scenario). As the experimental fields, we selected two touristic areas located in Germany and Japan, which have completely different conditions. We evaluated the emotion estimation model through a 3-class classification task (positive, neutral, negative) using UAR score as a metric, and achieved up to 0.50 of UAR score. Then, we evaluated the satisfaction estimation model through a seven-level regression task (0: fully unsatisfied – 6: fully satisfied) using MAE as a metric, and achieved up to 1.08 of MAE.

For the second challenge, in Chapter 4, we presented the design and implementation of a novel participatory sensing platform, named ParmoSense, that allows them to construct, customize, operate *Human-in-the-loop sensing* systems easily and quickly regardless of their ICT skills. To achieve this, we employ two features:

*modularization of functions* and *scenario-based participatory sensing system description*. We provide various functions essential for participatory sensing systems such as sensing functions, motivating functions for participants, and processing functions for collected data, and allow organizers to combine these modularized functions freely through a GUI web application. We call a combination of these modularized functions a *scenario*. Once a scenario has been created, participants can download it onto the ParmoSense client application and run it without doing any further setup or processing tasks. Thus, participants can contribute to many different sensing tasks without installing multiple-applications or performing complicated tasks which require technical skills. Through 17 case studies over two years, we confirmed that ParmoSense could be compatible flexibly with various sensing tasks, motivating methods, and data processing. Moreover, ParmoSense can suppress the burden of organizers and participants in system operation, and we found that it has higher cost performance than any conventional platforms.

Overall, our study has contributed the following academic knowledge. First, we have shown the potential of the *human-in-the-loop sensing* framework for gathering objective and subjective contexts dispersed in vast urban spaces. The subjective contexts understanding, which is largely unimplemented in existing urban environment systems, will help us to provide various systems in considered with context-awareness in a deeper level. Moreover, the correction of sensor characteristics will strongly support the crowdsourcing-based sensing approach, which utilizes low-quality sensors such as smartphone embedded sensors. Second, we have also explored that how can we apply our work to our daily life. Our proposed participatory sensing platform has shown a possibility that can be used in existing local citizen communities, especially civic tech communities for different purposes. The mutual linkage will help to make the system sustainable, and enhance the activities of a community.

## 5.2 Discussion and Future Perspective

### 5.2.1 Privacy Considerations

The *human-in-the-loop sensing* framework involves general people in the urban sensing system. This characteristic often causes privacy concern. For example, the strangers might be slipped into pictures or recordings when people collect data. To implement our framework in the real-world, we also should address to solve this issue.

The first idea is using alternative sensors that don't collect the private information. In this dissertation, we used the ideal devices (e.g., eye-tracking headset, camera) for building our proposed system, and confirmed the baseline performance. However, there is a possibility that such devices can be replaced with other devices. The second idea is excluding privacy information by processing data in edge-side. It deals privacy concern by delegating the pre-processing part to user's devices, and uploading data which doesn't include any private information.

As future work, we will consider incorporating these ideas to our system with keeping (or improving) its performance.

### 5.2.2 Knowledge Reusability

In this dissertation, we proposed two different sensing technology under the *human-in-the-loop sensing* framework. Here, we discuss the reusability of each knowledge.

In Chapter 2, we proposed the methodology to correct differences of sensor characteristics, especially the light sensor of smartphones, for the NightRoadScanner system. In fact, not only the light sensor but also other sensors have the same problem. Our method is conceptually based on a comparison between the correction standard device and correction target devices. It suggests that we can apply the NightRoadScanner knowledge to correct other sensors. For example, in EmoTour (Chapter 3), we use eye-tracking headset and head/body movement sensor under the calibrated condition, however, it consumes much time and labor. If we could apply the correction method into EmoTour, its realizable may

be improved.

In Chapter 3, we provided the methodology to collect geo-linked psychological data based on observation of user unconscious actions, named EmoTour. In this dissertation, we used the corrected data for understanding user satisfaction of sightseeing. By adding new labels, the target domain can be flexibly changed. For example, the ultimate purpose of NightRoadScanner (Chapter 2) is providing safe route navigation by assessing the safety level of sidewalks at night. To realize more reliable navigation algorithms, the feeling of users that can be collected using the EmoTour knowledge may be useful.

Additionally, we of course can apply knowledge found in this dissertation into other domains. As the future work, we will explore its possibility.

### **5.2.3 Method Generalizability**

In this dissertation, we have used special equipments which are not sold in general consumer markets, especially for EmoTour. To realize our proposed method in real-world conditions, we should consider using consumer devices, i.e., we have to generalize our proposed methodology. For example, the head/body movements can be collected using a combination of glasses-type wearables (e.g., GLASS [102]) and watch-type wearables (e.g., Fitbit [103], Google Smartwatches [104]) instead of SenStick.

Also, the generalization of the methodology will help to integrate knowledge from both NightRoadScanner and EmoTour into ParmoSense. In this dissertation, we have used ParmoSense elements partially in each system, but not completely integrated. We'll effort for this integration and conduct large-scale experiment with general people, as the future work.

### **5.2.4 Considerations of Personality and Preference**

In this dissertation, especially in Chapter 2 and Chapter 3, we provided the methods for estimating objective/subjective contexts in the urban space that can be utilized to make people's daily life more comfortable. To utilize these estimated contexts in the real-world systems such as a recommendation system, the personality and/or preferences of the user are also essential factors because the

user's decision-making is greatly susceptible to them. Thus, the discussion for connecting personality/preferences and our achievement provided in this dissertation should help to apply our knowledge in the real-world services.

As the future work, we will consider the routing algorithm which can provide an optimal route for each user in consideration of not only safety information estimated by NightRoadScanner system, but also user personality/preference information. Also, we will work on connecting EmoTour system and tourist guidance systems by collecting personality/preference information of tourist as an additional label.

# Acknowledgements

Upon the completion of this dissertation, I would like to express the deepest appreciation to important people, without whom this dissertation would not have been possible.

...

I would like to extend my deepest gratitude to Prof. Keiichi Yasumoto, who is my supervisor, for his guidance and encouragement throughout my research period. I learned a lot from him including academic thinking, technical writing, and how to progress research projects. Also, he gave me a chance to learn about organizing the academic conferences/workshops. His advice will be greatly helpful in my future academic activity.

I would also like to thank my co-supervisor, Prof. Satoshi Nakamura for reviewing this dissertation and providing feedback for improvement. He gave me a chance to study Ulm University, Germany, and kindly supported my stay.

I cannot begin to express my gratitude to Prof. Wolfgang Minker, who supervises me about the important research topic of this dissertation during I was studying at Ulm University, Germany. I learned a lot from him and his laboratory's members about the spoken dialogue systems, technical English writing, and others. Furthermore, I could spend a greatly meaningful time in Germany because he has always cared for me.

I must also greatly thankful to Associate Prof. Yutaka Arakawa, for his helpful suggestion and interesting idea from multidisciplinary research area. I learned a lot from him about not only academic domain but also startup or business domain

throughout many research projects. For example, technical writing, and how to progress research projects, additionally, brainstorming to conceive a new idea, the fusion thinking of academic and business.

I am also grateful to Associate Prof. Ismail Arai, for giving me the opportunity of starting my life as a researcher. He supervised me when I was studying at Akashi National College of Technology, seven years ago. I learned the basics of academic research from him, and it is still alive and always support me. He had encouraged me to go outside of college, and this experience motivated me to do not only academic research but also other activities.

I also wish to thank Assistant Prof. Hirohiko Suwa for giving me a lot of advice about statistical analysis, and co-writing papers. His suggestion always helped me to proceed with my research work, especially, analysis of case studies and its survey.

I would also like to thank Assistant Prof. Manato Fujimoto for giving me advice about the latest trends of technical paper and research proposal writing, and valuable discussion about research idea. He guided me in the choice of my career.

I very much appreciate Contract Assistant Prof. Teruhiro Mizumoto for the meaningful discussion about research idea and research proposal writing. These experiences will help me in future academic activity.

...

I am also much grateful to the laboratory's secretaries, Ms. Megumi Kanaoka, Ms. Nao Yamauchi, and Ms. Azusa Minami, who support my study and work in the laboratory. I would also like to thank all the members of the Ubiquitous Computing Systems Laboratory, especially Yugo Nakamura, Shogo Kawanaka for their friendship and support.

I would also like to thank all the members of the Institute of Communications Engineering at Ulm University, especially, institute's secretaries, Ms. Fe Hägele

and Ms. Michaela Baumann; and senior researcher, Dr. Werner Teich, who support me during my stay in Germany. Also, I am grateful to all the Ph.D. students, especially, Dmitrii Fedotov and Oleg Akhtiamov for their great friendship and support. Ich möchte mich recht herzlich bedanken! Я хотел бы поблагодарить вас!

...

I would like to thank my parents for their kind support. Finally, I would like to extend my deepest appreciation to my wife, Yuna Matsuda.

...

This work is partly supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers: [16J09670](#) – Research Fellowships for Young Scientists (DC1) and [16H01721](#) – Grant-in-Aid for Scientific Research (A); the Japan Science and Technology Agency (JST) Presto; and RIKEN, Center for Advanced Intelligence Project (AIP).

# References

- [1] GSMA. *The Mobile Economy 2018*. <https://www.gsma.com/mobileeconomy/>.
- [2] Mark Weiser. The computer for the 21st century. *Scientific american*, Vol. 265, No. 3, pp. 94–105, JSTOR, 1991.
- [3] Tomohiro Nishimura, Takamasa Higuchi, Hirozumi Yamaguchi, and Teruo Higashino. Detecting smoothness of pedestrian flows by participatory sensing with mobile phones. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, ISWC '14, pp. 15–18, ACM, 2014. doi: 10.1145/2634317.2642869.
- [4] Shigeya Morishita, Shogo Maenaka, Nagata Daichi, Morihiko Tamai, Keiichi Yasumoto, Toshinobu Fukukura, and Keita Sato. SakuraSensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UBIComp '15, pp. 695–705. ACM, 2015. doi: 10.1145/2750858.2804273.
- [5] Mani Srivastava, Tarek Abdelzaher, and Boleslaw Szymanski. Human-centric sensing. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, Vol. 370, No. 1958, pp. 176–197, The Royal Society, 2012. doi: 10.1098/rsta.2011.0244.
- [6] Charu C. Aggarwal and Tarek Abdelzaher. Social sensing. In *Managing and mining sensor data*, pp. 237–297. Springer, 2013. doi: 10.1007/978-1-4614-6309-2\_9.

- [7] Jeffrey A. Burke, D. Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, and Mani B. Srivastava. Participatory sensing. *Center for Embedded Network Sensing*, 2006.
- [8] Nicolas Maisonneuve, Matthias Stevens, and Bartek Ochab. Participatory noise pollution monitoring using mobile phones. *Information Polity*, Vol. 15, No. 1, 2, pp. 51–71, ACM, 2010. doi: 10.3233/IP-2010-0200.
- [9] Eiman Kanjo. NoiseSPY: a real-time mobile phone platform for urban noise monitoring and mapping. *Mobile Networks and Applications*, Vol. 15, No. 4, pp. 562–574, Springer US, 2010. doi: 10.1007/s11036-009-0217-y.
- [10] OpenSignal, Inc. OpenSignal. <http://opensignal.com/>. (accessed 31 Jan. 2019).
- [11] Diego Mendez, Alfredo J Perez, Miguel Labrador, Juan Jose Marron, et al. P-Sense: A participatory sensing system for air pollution monitoring and control. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops*, PerCom '11 Workshops, pp. 344–347. IEEE, 2011. doi: 10.1109/PERCOMW.2011.5766902.
- [12] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. U-Air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '13, pp. 1436–1444. ACM, 2013. doi: 10.1145/2487575.2488188.
- [13] Motorola Solutions, Inc. CrimeReports.com. <https://www.crimereports.com/>. (accessed 31 Jan. 2019).
- [14] Japan Security Systems Association. Guide of security illumination, Vol. 5. p. 2, 2013. (in Japanese).
- [15] Yuki Matsuda, Ismail Arai, Yutaka Arakawa, and Keiichi Yasumoto. A correction method for individual differences among smartphones' light sensor to measure streetlamps' illuminance for sidewalks at night. *IPSSJ Journal*, Vol. 57, No. 2, pp. 520–531, IPSJ, 2016. (in Japanese).

- [16] Yoshihito Yamada, Masayuki Aoki, Hiroki Watano, and Masao Miura. Understanding of residents' awareness by questionnaire survey and measurement of night illuminance after installation of streetlights : Support of further activation of improvement activities of living environment in the area where new streetlights were installed. part 1. *Summaries of technical papers of annual meeting*, Vol. 2013, pp. 1139–1140, Architectural Institute of Japan, 2013. (in Japanese).
- [17] Go Takano, Akane Okui, Yoshihito Yamada, and Masao Miura. Extraction of prior improvement point and planning of night illuminance improvement plan through questionnaire survey and night illuminance measurement : The support of night illuminance improvement activities by residents' initiatives in the area with farmlands and residential district part 2. *Summaries of technical papers of annual meeting*, pp. 1155–1156, Architectural Institute of Japan, 2014. (in Japanese).
- [18] Japanese Industrial Standards Committee. Illuminance measurements for lighting installations. *JIS C 7612:1985*, 1985. (in Japanese).
- [19] Japan Security Systems Association. Standard for levels of illumination on crime prevention lighting. *SES E1901-4*, 2015. (in Japanese).
- [20] Japanese Industrial Standards Committee. Lighting for roads. *JIS Z 9111:1988*, 1988. (in Japanese).
- [21] Yuki Matsuda and Ismail Arai. A safety assessment system for sidewalks at night utilizing smartphones' light sensors. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, pp. 115–118, ACM, 2014. doi: 10.1145/2638728.2638797.
- [22] Yuki Matsuda and Ismail Arai. An experiment of a streetlamp classifying and a vertical illuminance assessing utilizing smartphones' light sensors. *3rd International Symposium on Technology for Sustainability*, No. 264, 2013.

- [23] Yuki Matsuda and Ismail Arai. Comprehensive gathering system for street-lamps brightness utilizing collective smartphone light sensor data. *IPSSJ Journal*, Vol. 55, No. 2, pp. 750–760, IPSJ, 2013. (in Japanese).
- [24] Yuki Matsuda and Ismail Arai. Development of safety assessment system of night sidewalk in consideration of the incident light characteristics utilizing collective smartphone light sensor data. In *Proceedings of Multimedia, Distributed, Cooperative, and Mobile Symposium*, DICOMO '14, pp. 966–974. IPSJ, 2014. (in Japanese).
- [25] Joaquín Alegre and Jaume Garau. Tourist satisfaction and dissatisfaction. *Annals of Tourism Research*, Vol. 37, No. 1, pp. 52–73, Elsevier, 2010. doi: 10.1016/j.annals.2009.07.001.
- [26] Ching F. Chen and Fu S. Chen. Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists. *Tourism Management*, Vol. 31, No. 1, pp. 29–35, Elsevier, 2010. doi: 10.1016/j.tourman.2009.02.008.
- [27] TripAdvisor. <http://www.tripadvisor.com/>. (accessed 31 Jan. 2019).
- [28] Yelp. <https://www.yelp.com/>. (accessed 31 Jan. 2019).
- [29] Amazon. <https://www.amazon.com/>. (accessed 31 Jan. 2019).
- [30] Kun Han, Dong Yu, and Ivan Tashev. Speech emotion recognition using deep neural network and extreme learning machine. In *15th Annual Conference of the International Speech Communication Association*, InterSpeech '14, 2014.
- [31] Heysem Kaya, Alexey A. Karpov, and Albert Ali Salah. Robust acoustic emotion recognition based on cascaded normalization and extreme learning machines. In *13th International Symposium on Neural Networks*, ISNN '16, pp. 115–123. Springer Cham, 2016. doi: 10.1007/978-3-319-40663-3\_14.
- [32] Wei Y. Qu, Dong Y. Huang, Weisi Lin, Haizhou Li, and Minghui Dong. Mobile acoustic emotion recognition. In *2016*

- IEEE Region 10 Conference, TENCON '16*, pp. 170–174. IEEE, 2016. doi: 10.1109/TENCON.2016.7847983.
- [33] Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz J. Rak. Emotion recognition using facial expressions. *Procedia Computer Science*, Vol. 108, pp. 1175–1184, Elsevier, 2017. doi: 10.1016/j.procs.2017.05.025.
- [34] Panagiotis Tzirakis, George Trigeorgis, Mihalis A. Nicolaou, Björn W. Schuller, and Stefanos Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 11, No. 8, pp. 1301–1309, IEEE, 2017. doi: 10.1109/JSTSP.2017.2764438.
- [35] Arulkumar Subramaniam, Vismay Patel, Ashish Mishra, Prashanth Balasubramanian, and Anurag Mittal. Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features. In *Computer Vision - ECCV 2016 Workshops, ECCV '16*, pp. 337–348. Springer Cham, 2016. doi: 10.1007/978-3-319-49409-8.
- [36] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition, FG '13*, pp. 1–8. IEEE, 2013. doi: 10.1109/FG.2013.6553805.
- [37] Abhinav Dhall, Roland Goecke, Shreya Ghosh, Jyoti Joshi, Jesse Hoey, and Tom Gedeon. From individual to group-level emotion recognition: Emotiw 5.0. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction, ICMI '17*, pp. 524–528, ACM, 2017. doi: 10.1145/3136755.3143004.
- [38] Maxim Sidorov and Wolfgang Minker. Emotion recognition in real-world conditions with acoustic and visual features. In *Proceedings of the 16th International Conference on Multimodal Interaction, ICMI '14*, pp. 521–524, ACM, 2014. doi: 10.1145/2663204.2666279.

- [39] Ping Hu, Dongqi Cai, Shandong Wang, Anbang Yao, and Yurong Chen. Learning supervised scoring ensemble for emotion recognition in the wild. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, ICMI '17, pp. 553–560, ACM, 2017. doi: 10.1145/3136755.3143009.
- [40] Shams Shapsough, Ahmed Hesham, Youssef Elkhorary, Imran A. Zuakernan, and Fadi Aloul. Emotion recognition using mobile phones. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services*, Healthcom '16, pp. 1–6. IEEE, 2016. doi: 10.1109/HealthCom.2016.7749470.
- [41] Bernd Resch, Anja Summa, Günther Sagl, Peter Zeile, and Jan-Philipp Exner. Urban Emotions – geo-semantic emotion extraction from technical sensors, human sensors and crowdsourced data. In *Progress in Location-Based Services 2014*, pp. 199–212. Springer Cham, 2014. doi: 10.1007/978-3-319-11879-6\_14.
- [42] Panagiotis C. Petrantonakis and Leontios J. Hadjileontiadis. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on Affective Computing*, Vol. 1, No. 2, pp. 81–97, IEEE, 2010. doi: 10.1109/T-AFFC.2010.7.
- [43] Yuan-Pin Lin, Chi-Hong Wang, Tien-Lin Wu, Shyh-Kang Jeng, and Jyh-Horng Chen. Eeg-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 489–492, 2009. doi: 10.1109/ICASSP.2009.4959627.
- [44] Fabien Ringeval, Florian Eyben, Eleni Kroupi, Anil Yuce, Jean P. Thiran, Touradj Ebrahimi, Denis Lalanne, and Björn W. Schuller. Prediction of asynchronous dimensional emotion ratings from audiovisual and physiological data. *Pattern Recognition Letters*, Vol. 66, pp. 22–30, Elsevier, 2015. doi: 10.1016/j.patrec.2014.11.007.
- [45] Tuka W. AlHanai and Mohammad M. Ghassemi. Predicting latent narrative mood using audio and physiologic data. In *Proceedings of the Thirty-*

*First AAAI Conference on Artificial Intelligence*, AAAI '17, pp. 948–954. AAAI, 2017.

- [46] Wei L. Zheng, Bo N. Dong, and Bao L. Lu. Multimodal emotion recognition using EEG and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC '14*, pp. 5040–5043. IEEE, 2014. doi: 10.1109/EMBC.2014.6944757.
- [47] Mohammad Soleymani, Maja Pantic, and Thierry Pun. Multimodal emotion recognition in response to videos. *IEEE transactions on affective computing*, Vol. 3, No. 2, pp. 211–223, IEEE, 2012. doi: 10.1109/T-AFFC.2011.37.
- [48] Mohammad Soleymani, Sadjad Asghari-Esfeden, Yun Fu, and Maja Pantic. Analysis of eeg signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, Vol. 7, No. 1, pp. 17–28, 2016. doi: 10.1109/TAFFC.2015.2436926.
- [49] Eiman Kanjo, Eman M. G. Younis, and Nasser Sherkat. Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach. *Information Fusion*, Vol. 40, pp. 18–31, Elsevier, 2018. doi: 10.1016/j.inffus.2017.05.005.
- [50] Jumpei Yamamoto, Mizuki Kawazoe, Jin Nakazawa, Kazunori Takashio, and Hideyuki Tokuda. MOLMOD: analysis of feelings based on vital information for mood acquisition. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Vol. 4 of MobileHCI '09, pp. 1–4, ACM, 2009.
- [51] Yuki Matsuda, Dmitrii Fedotov, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, and Wolfgang Minker. EmoTour: Estimating emotion and satisfaction of users based on behavioral cues and audiovisual data. *Sensors*, Vol. 18, No. 11, MDPI, 2018. doi: 10.3390/s18113978.
- [52] Ekaterina Balandina, Sergey Balandin, Yevgeni Koucheryavy, and Dmitry Mouromtsev. IoT use cases in healthcare and tourism. In *2015 IEEE*

- 17th Conference on Business Informatics*, Vol. 2, pp. 37–44. IEEE, 2015. doi: 10.1109/CBI.2015.16.
- [53] Masato Hidaka, Yuki Matsuda, Shogo Kawanaka, Yugo Nakamura, Manato Fujimoto, Yutaka Arakawa, and Keiichi Yasumoto. A system for collecting and curating sightseeing information toward satisfactory tour plan creation. In *The Second International Workshop on Smart Sensing Systems, IWSSS '17*, 2017.
- [54] Chung H. Wu, Jen C. Lin, and Wen L. Wei. Survey on audiovisual emotion recognition: databases, features, and data fusion strategies. *APSIPA transactions on signal and information processing*, Vol. 3, p. E12, Cambridge University Press, 2014. doi: 10.1017/ATSIP.2014.11.
- [55] Dmitrii Fedotov, Yuki Matsuda, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, and Wolfgang Minker. Towards estimating emotions and satisfaction level of tourist based on eye gaze and head movement. In *2018 IEEE International Conference on Smart Computing, SMARTCOMP*, pp. 399–404. IEEE, 2018. doi: 10.1109/SMARTCOMP.2018.00036.
- [56] Moritz Kassner, William Patera, and Andreas Bulling. Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp '14 Adjunct*, pp. 1151–1160. ACM, 2014. doi: 10.1145/2638728.2641695.
- [57] Yugo Nakamura, Yutaka Arakawa, Takuya Kanehira, Masashi Fujiwara, and Keiichi Yasumoto. SenStick: Comprehensive sensing platform with an ultra tiny all-in-one sensor board for iot research. *Journal of Sensors*, Vol. 2017, Hindawi, 2017. doi: 10.1155/2017/6308302.
- [58] James A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, Vol. 39, No. 6, pp. 1161–1178, APA, 1980. doi: 10.1037/h0077714.
- [59] Georgios Paltoglou and Michael Thelwall. Seeing stars of valence and

- arousal in blog posts. *IEEE Transactions on Affective Computing*, Vol. 4, No. 1, pp. 116–123, IEEE, 2013. doi: 10.1109/T-AFFC.2012.36.
- [60] Hong Ying, C. Silex, A. Schnitzer, S. Leonhardt, and M. Schiek. Automatic step detection in the accelerometer signal. In *4th International Workshop on Wearable and Implantable Body Sensor Networks*, BSN '07, pp. 80–85. Springer Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-70994-7\_14.
- [61] Florian Eyben, Martin Wöllmer, and Björn W. Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, pp. 1459–1462. ACM, 2010. doi: 10.1145/1873951.1874246.
- [62] Björn W. Schuller, Stefan Steidl, Anton Batliner, Julien Epps, Florian Eyben, Fabien Ringeval, Erik Marchi, and Yue Zhang. The interspeech 2014 computational paralinguistics challenge: Cognitive & physical load. In *Fifteenth Annual Conference of the International Speech Communication Association*, InterSpeech '14, pp. 427–431, 2014.
- [63] Tadas Baltrušaitis. OpenFace. <https://github.com/TadasBaltrusaitis/OpenFace>. (accessed 1 Jan. 2019).
- [64] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition*, FG '18, pp. 59–66. IEEE, 2018. doi: 10.1109/FG.2018.00019.
- [65] Paul Ekman and Wallace V. Friesen. *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.
- [66] Paul Ekman and Erika L. Rosenberg. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.
- [67] Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schroder. The SEMAINE Database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent.

*IEEE Transactions on Affective Computing*, Vol. 3, No. 1, pp. 5–17, IEEE, 2012. doi: 10.1109/T-AFFC.2011.20.

- [68] Angeliki Metallinou, Zhaojun Yang, Chi-Chun Lee, Carlos Busso, Sharon Carnicke, and Shrikanth Narayanan. The USC CreativeIT Database of Multimodal Dyadic Interactions: From speech and full body motion capture to continuous emotional annotations. *Lang. Resour. Eval.*, Vol. 50, No. 3, pp. 497–521, Springer Netherlands, 2016. doi: 10.1007/s10579-015-9300-0.
- [69] Olga Perepelkina, Evdokia Kazimirova, and Maria Konstantinova. RAMAS: Russian multimodal corpus of dyadic interaction for affective computing. In *International Conference on Speech and Computer*, pp. 501–510. Springer Cham, 2018. doi: 10.1007/978-3-319-99579-3\_52.
- [70] Dmitrii Fedotov, Denis Ivanko, Maxim Sidorov, and Wolfgang Minker. Contextual dependencies in time-continuous multidimensional affect recognition. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC '18*, pp. 1220–1224. ELRA, 2018.
- [71] Tijmen Tieleman and Geoffrey Hinton. *Lecture 6.5-RMSProp, COURSE-ERA: Neural networks for machine learning*. University of Toronto, Technical Report. 2012.
- [72] Jennifer Tehan Stanley, Xin Zhang, Helene H Fung, and Derek M Isaacowitz. Cultural differences in gaze and emotion recognition: Americans contrast more than chinese. *Emotion*, Vol. 13, No. 1, pp. 36–46, APA, 2013. doi: 10.1037/a0029209.
- [73] Louisa Pragst, Stefan Ultes, Matthias Kraus, and Wolfgang Minker. Adaptive dialogue management in the KRISTINA project for multicultural health care applications. In *Proceedings of the 19th Workshop on the Semantics and Pragmatics of Dialogue, SEMDIAL '15*, pp. 202–203, 2015.
- [74] Juliana Miehle, Wolfgang Minker, and Stefan Ultes. What causes the differences in communication styles? a multicultural study on directness and elaborateness. In *Proceedings of the Eleventh International Conference*

*on Language Resources and Evaluation*, LREC '18, pp. 3961–3968. ELRA, 2018.

- [75] Shinichiro Kanoh, Susumu Ichi-nohe, Shunsuke Shioya, Kazutaka Inoue, and Ryota Kawashima. Development of an eyewear to measure eye and body movements. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, EMBC '15, pp. 2267–2270. IEEE, 2015. doi: 10.1109/EMBC.2015.7318844.
- [76] Denzil Ferreira, Vassilis Kostakos, and Anind K. Dey. AWARE: Mobile context instrumentation framework. *Frontiers in ICT*, Vol. 2, No. 6, pp. 1–9, Frontiers, 2015.
- [77] Moo-Ryong Ra, Bin Liu, Tom F. La Porta, and Ramesh Govindan. Medusa: A programming framework for crowd-sensing applications. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pp. 337–350, ACM, 2012. doi: 10.1145/2307636.2307668.
- [78] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, Vol. 7, No. 6, pp. 643–659, Elsevier, 2011. doi: 10.1016/j.pmcj.2011.09.004.
- [79] Mina Sakamura, Tomotaka Ito, Hideyuki Tokuda, Takuro Yonezawa, and Jin Nakazawa. MinaQn: Web-based participatory sensing platform for citizen-centric urban development. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, UbiComp/ISWC'15 Adjunct, pp. 1607–1614, ACM, 2015. doi: 10.1145/2800835.2801632.
- [80] HongSuda Tangmunarunkit, Cheng-Kang Hsieh, Brent Longstaff, S Nolen, John Jenkins, Cameron Ketcham, Joshua Selsky, Faisal Alquaddoomi, Dony George, Jinha Kang, Z. Khalapyan, J. Ooms, Nithya Ramanathan,

- and Deborah L Estrin. Ohmage: A general and extensible end-to-end participatory sensing platform. *ACM Transactions on Intelligent Systems and Technology*, Vol. 6, No. 3, pp. 38:1–38:21, ACM, 2015. doi: 10.1145/2717318.
- [81] Waylon Brunette, Mitchell Sundt, Nicola Dell, Rohit Chaudhri, Nathan Breit, and Gaetano Borriello. Open Data Kit 2.0: Expanding and refining information services for developing regions. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, HotMobile '13, pp. 10:1–10:6, ACM, 2013. doi: 10.1145/2444776.2444790.
- [82] Yutaka Arakawa and Yuki Matsuda. Gamification mechanism for enhancing a participatory urban sensing: survey and practical results. *Journal of Information Processing*, Vol. 24, No. 1, pp. 31–38, IPSJ, 2016. doi: 10.2197/ipsjjip.24.31.
- [83] Gamification Co. Gamification pitfalls: Badge fatigue and loyalty backlash. <http://www.gamification.co/2012/09/12/gamification-pitfalls-badge-fatigue-and-loyalty-backlash/>, 2012.
- [84] Francesco Restuccia, Sajal K. Das, and Jamie Payton. Incentive mechanisms for participatory sensing: Survey and research challenges. *ACM Transactions on Sensor Networks*, Vol. 12, No. 2, 2016. doi: 10.1145/2888398.
- [85] Luis G Jaimes, Idalides Vergara-Laurens, and Miguel A Labrador. A location-based incentive mechanism for participatory sensing systems with budget constraints. In *2012 IEEE International Conference on Pervasive Computing and Communications*, PerCom '12, pp. 103–108. IEEE, 2012. doi: 10.1109/PerCom.2012.6199855.
- [86] Hui Gao, Chi Harold Liu, Wendong Wang, Jianxin Zhao, Zheng Song, Xin Su, Jon Crowcroft, and Kin K. Leung. A Survey of Incentive Mechanisms for Participatory Sensing. *IEEE Communications Surveys & Tutorials*, Vol. 17, No. 2, pp. 918–943, IEEE, 2015. doi: 10.1109/COMST.2014.2387836.

- [87] Robert Ighodaro Ogie. Adopting incentive mechanisms for large-scale participation in mobile crowdsensing: from literature review to a conceptual framework. *Human-centric Computing and Information Sciences*, Vol. 6, No. 1, pp. 80:1–80:31, Springer Berlin Heidelberg, 2016. doi: 10.1186/s13673-016-0080-3.
- [88] Gabe Zichermann and Christopher Cunningham. *Gamification by design: Implementing game mechanics in web and mobile apps*. O’Reilly Media, Inc., 2011.
- [89] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek ’11, pp. 9–15. ACM, 2011. doi: 10.1145/2181037.2181040.
- [90] Fabian Groh. Gamification: State of the art definition and utilization. *Institute of Media Informatics Ulm University*, Vol. 39, 2012.
- [91] Yoshitaka Ueyama, Morihiko Tamai, Yutaka Arakawa, and Keiichi Yasumoto. Gamification-based incentive mechanism for participatory sensing. In *2014 IEEE International Conference on Pervasive Computing and Communications Workshops*, PerCom ’14 Workshops, pp. 98–103. IEEE, 2014. doi: 10.1109/PerComW.2014.6815172.
- [92] MIT Media Lab. Funf: Open sensing framework. <http://funf.org/>. (accessed 31 Jan. 2019).
- [93] OASIS. MQTT.org. <http://mqtt.org/>. (accessed 31 Jan. 2019).
- [94] OpenStreetMap. <http://www.openstreetmap.org/>. (accessed 31 Jan. 2019).
- [95] Google LLC. Google Earth. <https://earth.google.com/>. (accessed 31 Jan. 2019).
- [96] Secium. <https://cesiumjs.org/>. (accessed 31 Jan. 2019).

- [97] Desislava Hristova, Giovanni Quattrone, Afra Mashhadi, and Licia Capra. The life of the party: Impact of social mapping in OpenStreetMap. In *The International AAAI Conference on Weblogs and Social Media*, ICWSM '13. AAAI, 2013.
- [98] Peter Mooney, Marco Minghini, and Frances Stanley-Jones. Observations on an openstreetmap mapping party organised as a social event during an open source GIS conference. *International Journal of Spatial Data Infrastructures Research (IJS DIR)*, Vol. 10, pp. 138–150, 2015. doi: 10.2902/1725-0463.2015.10.art7.
- [99] Mordechai Haklay and Patrick Weber. OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing*, Vol. 7, No. 4, pp. 12–18, IEEE, 2008. doi: 10.1109/MPRV.2008.80.
- [100] FixMyStreet.org. FixMyStreet Platform. <http://fixmystreet.org/>. (accessed 31 Jan. 2019).
- [101] FaithCreates, Inc. Rally. <https://rallyapp.jp/>. (accessed 31 Jan. 2019).
- [102] X DEVELOPMENT LLC. Glass. <https://x.company/glass/>. (accessed 31 Jan. 2019).
- [103] Fitbit, Inc. fitbit. <https://www.fitbit.com/>. (accessed 31 Jan. 2019).
- [104] Google LLC. Wear OS by Google Smartwatches. <https://wearos.google.com/>. (accessed 31 Jan. 2019).

# Appendix

## A. Derivation of Illuminance Model at Walking

This section provides the method for deriving illuminance model at walking. A detailed positional relationship between a streetlamp and a smartphone embedded light sensor is shown in Figure A.1. Assuming the streetlamp has a diffusive point light source with even luminosity in any direction, Illuminance value  $\hat{E}_s$  when the user holds the light sensor at right angles to the light source can be calculated with Equation (A.1) according to the inverse-square law of light.

$$\hat{E}_s = E_h \cdot \cos^2 \theta_l \quad (\text{A.1})$$

Here, in the case when the user holds the light sensor with the angle  $\theta_s$ , the surface that obtains the illuminance vertically from the light source tilts at angle  $\theta_x$  calculated with Equation (A.2). The parameter  $\theta_{l+s}$  represents the sum of angle  $\theta_l$  and  $\theta_s$ .

$$\theta_x = |90 - (\theta_l + \theta_s)| = |90 - \theta_{l+s}| \quad (\text{A.2})$$

However, there is the blind spot  $\alpha_\theta$  in the light sensor of a smartphone as shown in Figure A.2. The angle  $\hat{\theta}_{l+s}$  in consideration of this blind spot  $\alpha_\theta$  can be derived with Equation (A.3).

$$\hat{\theta}_{l+s} = \begin{cases} 0 & (\theta_{l+s} \leq \alpha_\theta) \\ \theta_l + \theta_s - \alpha_\theta & (\theta_{l+s} > \alpha_\theta) \end{cases} \quad (\text{A.3})$$

The range of the angle  $\hat{\theta}_{l+s}$  becomes  $0 \leq \hat{\theta}_{l+s} \leq 60$ . Hence, we change this range  $0 \leq \alpha_{\text{ratio}} \cdot \hat{\theta}_{l+s} \leq 90$  using the parameter  $\alpha_{\text{ratio}}$ .  $\alpha_{\text{ratio}}$  is derived with Equation (A.4).

$$\alpha_{\text{ratio}} = \frac{90}{90 - \alpha_\theta} = 1.5 \quad (\text{A.4})$$

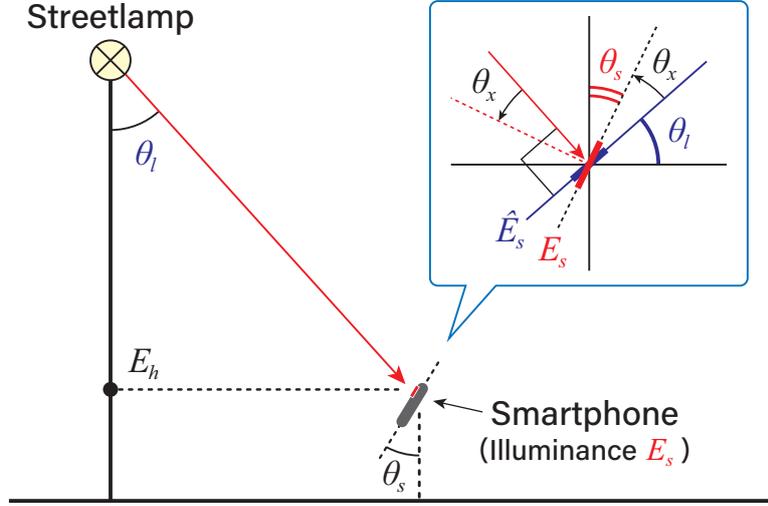


Figure A.1. A positional relationship between a streetlamp and a smartphone embedded light sensor.

The illuminance value  $\bar{E}_s$ , when the user holds the light sensor at angle  $\theta_s$  to the light source, can be derived with Equation (A.5) by using values derived through the above process.

$$\begin{aligned}
 \bar{E}_s &= \hat{E}_s \cdot \cos |90 - \alpha_{\text{ratio}} \cdot (\theta_l + \theta_s - \alpha_\theta)| \\
 &= \hat{E}_s \cdot \sin (\alpha_{\text{ratio}} \cdot (\theta_l + \theta_s - \alpha_\theta)) \\
 &= E_h \cdot \cos^2 \theta \cdot \sin (\alpha_{\text{ratio}} \cdot (\theta_l + \theta_s - \alpha_\theta))
 \end{aligned} \tag{A.5}$$

The actual light source is not the diffusive point light source with the uniform luminous intensity. Hence, the luminous intensity in the direction at angle  $\theta$  attenuates in proportion to the cosine compared to the case of 0 degree. Taking this attenuation into consideration, the illuminance  $E_s$  actually measured by the smartphone is derived with Equation (A.6).

$$\begin{aligned}
 E_s &= \bar{E}_s \cdot \cos \theta_l \\
 &= E_h \cdot \cos^3 \theta_l \cdot \sin (\alpha_{\text{ratio}} \cdot (\theta_l + \theta_s - \alpha_\theta))
 \end{aligned} \tag{A.6}$$

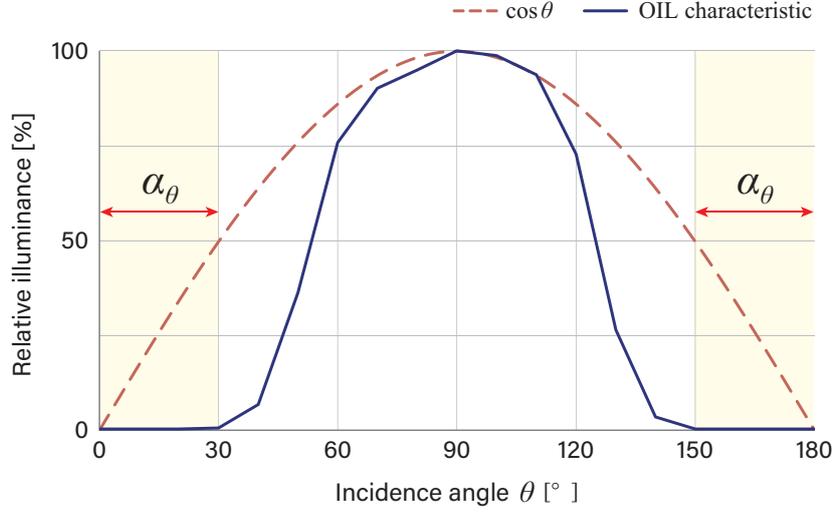


Figure A.2. OIL characteristic of Galaxy Nexus.

## B. Correction of Smartphone Holding Angle

This section provides the method of correcting the smartphone holding angle. A positional relationship between a streetlamp and a smartphone embedded light sensor is shown in Figure B.1.

Assuming the angle between the light source and the light sensor of a smartphone becomes  $\theta_l$  when the user holds smartphone at the angle  $\theta_s$ , The incidence angle of light  $\hat{\theta}_l$  can be calculated with Equation (B.1).

$$\hat{\theta}_l = \theta_l - (90 - \theta_s) \quad (\text{B.1})$$

Here, we need to correct this angle to  $\theta_s = 90$ , and calculate illuminance values. The illuminance value  $E_s$ , which the smartphone holding angle is corrected, can be derived with Equation (B.2) by using measured illuminance value  $\hat{E}_s$ . in this condition, the OIL characteristic follow the Lambert's cosine law, which is a basic characteristic of light.

$$E_s = \hat{E}_s \times \frac{\cos(90 - \theta_l)}{\cos(90 - \hat{\theta}_l)} \quad (\text{B.2})$$

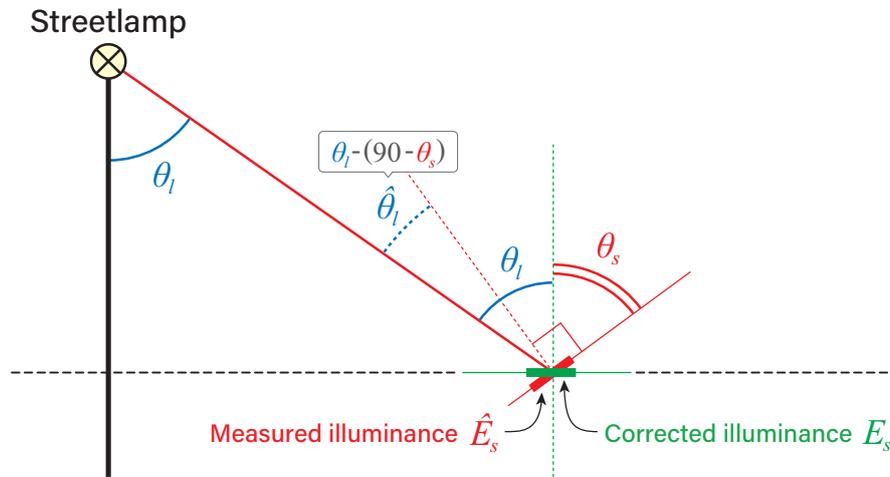


Figure B.1. The correction of the holding angle of a smartphone.

Also, if we assume that the illuminance incident angle characteristics follow the cosine to the 5th power ( $\cos^5$ ) obtained from Figure 2.6, the illuminance value  $E_s$ , which the smartphone holding angle is corrected, can be calculated with Equation (B.3).

$$E_s = \hat{E}_s \times \frac{\cos^5(90 - \theta_l)}{\cos^5(90 - \hat{\theta}_l)} \quad (\text{B.3})$$

## C. Sample Scenario of ParmoSense

---

```
1 {
2   "app_id": 20,
3   "app_title": "Sample scenario",
4   "app_description": "This is sample.",
5   "client": {
6     "sensing_func": {
7       "accelerometer": {
8         "enabled": true,
9         "interval": 200,
10        "description": "for step counting"
11      },
12      "gyroscope": { ... }, ...
13    },
14    "motivating_func": {
15      "ranking": {
16        "enabled": false,
17      },
18      "level": {
19        "enabled": true,
20        "function": 2,
21        "max": 100
22      },
23      "point": { ... }, ...
24    }
25  }
26  "dashboard": {
27    "editor": {
28      "enabled": false
29    },
30    "browser": {
31      "enabled": true,
32      "type": 0,
33      ...
34    }
35  }
36 }
```

---

Listing C.1. Sample of scenario in JSON format (not compressed)

# Glossary

---

## A

---

**Action Unit (AU)** is the specific movement (action) of individual or groups of facial muscles to describe facial expressions of human beings.

..... 46, 49, 131

**Android** is a mobile operating system developed by Google LLC. It is designed mainly for touchscreen mobile devices such as smartphones and tablets. In recent years, it is expanding the scope of supportable devices, e.g., wearable devices, vehicles, televisions. – <https://www.android.com/>

..... 43, 77, 134

**Application Programming Interface (API)** is a set of commands, functions, protocols, and objects for developing software or communicating with an external system. It makes developers easier to construct software applications without writing the code from scratch.

..... 78

---

## B

---

**Bluetooth Low Energy (BLE)** is a wireless personal area network technology, operating in the 2.4 GHz frequency band, designed for short burst data transmission. Its standard is defined by Bluetooth Special Interest Group (SIG). It is also marked as Bluetooth LE, Bluetooth Smart.

..... 80, 91, 133

**Butterworth filter** is a one of signal processing filter which is designed to have a frequency response as flat as possible in the passband.

..... 48

---

## C

---

**Chord diagram** is one of the diagrams which visualizes the inter-relationships between entities. Each node is placed along a circle, and the relationship between nodes shows as the bezier curves. The width of the curve represents the proportion of connection.

..... 92

**Civic technology (civic tech)** is the paradigm of the local community for promoting community development by combining information and communication technology (ICT) and civic cooperation which people work with government, universities, companies, etc.  
 ..... 8, 62, 99, 103

**Context-aware system** is a system that can gather information about surrounding environment, and analyze the situation (called *context*), and provide an appropriate service to users.  
 ..... 1, 3, 6, 9, 37, 38, 101

**Correction parameter** is the parameter for correcting illuminance transition in this dissertation (see Section 2.3.2). It is the distance differences of both positions when the illuminance value obtained by the correction standard/target device becomes the same value.  
 ..... 20, 21, 22, 28, 31, 32

**CreativeIT database** is a multimodal database which consists of 2.2 hours (31 sessions) of audio and motion capture recordings in English from 15 participants. It has been recorded by actors, coordinated by a director with expert qualification in Active Analysis introduced by Stanislavsky. – <https://sail.usc.edu/CreativeIT/>  
 ..... 49, 50

**Crowdsourcing** is the outsourcing method for obtaining needed services, ideas, or content through recruiting contributions from public people and especially from the online community.  
 ..... 4, 103

**CSV (Comma-Separated Values)** is a plain text file storing the table-like data which are delimited using a comma.  
 ..... 74, 82

**Cutoff frequency** is the frequency either above or below which the output of filter has fallen to one half ( $\approx 3$  dB) in the passband.  
 ..... 48

---

**E**

---

**Electrocardiogram (ECG)** is a noninvasive monitoring method to record the electrical rhythm of hearts, i.e., the electrical impulses of the heart muscle.  
 ..... 40

**Electrodermal activity (EDA)** is a measurement method to record the changes in conductivity produced in the skin due to increases in the activity of sweat glands. It is also

known as skin conductance, galvanic skin response (GSR).  
..... 40

**Electroencephalography (EEG)** is an electrophysiological noninvasive monitoring method to record the electrical activity of the brain. It measures voltage fluctuations resulting from ionic current within the neurons of the brain.  
..... 40

**Electrooculography (EOG)** is an electrophysiological monitoring method to measure the corneo-retinal potential between the front and the back of the human eye.  
..... 60

---

## F

---

**Facial action coding system (FACS)** is a coding system for taxonomizing human facial movements by their appearance on the face. In this system, AUs are used to describe human facial expressions.  
..... 49

**Feed-forward neural network** is the most general deep learning model. The signal propagates single direction from the input nodes to the output nodes through intermediate nodes, so that it named the *feedforward* neural network. It also known as multilayer perceptron (MLP).  
..... 51

---

## G

---

**Gamification** is one of the non-monetary incentive mechanisms applying the game theory concepts (the game elements and experience design) to non-game systems such as sensing systems, in order to engage and motivate people. It gives a kind of fun (e.g., point, ranking) as a virtual reward corresponding to the user's contribution.  
..... 67, 78, 81, 84, 136

**Global Positioning System (GPS)** is a satellite-based radio-navigation system owned by the United States. It provides users with positioning, navigation, and timing (PNT) services. – <https://www.gps.gov/>  
..... 12, 20, 43, 63, 80, 92, 133

**Graphical User Interface (GUI)** GUI is a form of user interface that allows users to operate the computer via graphical views, e.g., graphical icons and visual indicators, instead of text-based user interfaces such as CUI (character user interface). Most of recent modern operating systems on consumer computers provide users this interface.  
..... 71, 74, 85, 102

**Hubeny formula** is a method for calculating the distance between arbitrary two points on the earth from latitude and longitude data ( $d = \sqrt{(d_y R)^2 + (d_x N \cos \mu_y)^2}$ , the parameter  $d_x$  and  $d_y$  represents the longitude/latitude difference between two points respectively,  $\mu_y$  represents the average latitude of the two points,  $R$  represents the radius of curvature of the meridian, and  $N$  represents the transverse radius of curvature).

..... 20

**Human-centric sensing** is a sensing method proposed by Mani et al., which is based on the cooperation of human and machines by involving humans to the sensing process. It consists of two human involvement levels: *human as data sources* (*human as a data source* in this dissertation) and *human as sensor operators* (*human as a device carrier* in this dissertation), and *human as targets of sensing*.

..... 3

**Human-in-the-loop sensing** is an urban environmental sensing framework consisting of the two involvement levels: *human as a device carrier* and *human as a sensor*, which is defined in this dissertation.

..... vi, 4, 6, 8, 10, 15, 38, 62, 99, 101, 102, 103, 104

**Human involvement model** is a relationship structure of urban sensing approach organized in the viewpoint of human involvement in the system. It has four concepts of the involvement level: *Human as an end-consumer*, *human as a data source*, *human as a device carrier*, and *human as a sensor*.

..... vi, 2

**Human as a data source** is one of the concepts of human involvement model defined in this dissertation. In this concept, the urban sensing system employs a method based on diverting and analyzing a huge amount of SNS data posted by people who are not conscious of the contribution to the system. It enables to collect both objective and subjective contexts that can not be measured by sensors but can be recognized by the human.

..... 3, 132

**Human as a device carrier** is one of the concepts of human involvement model defined in this dissertation. In this concept, the urban sensing system employs a method based on requesting general people in the environment to contribute to sensing. By using sensors embedded in mobile devices owned by general people, it is possible to collect information equivalent to sensor networks without needing infrastructure. In addition, it is also possible to collect spatially or temporally insufficient data by daringly asking people.

..... vi, 3, 4, 5, 6, 8, 9, 10, 35, 101, 132

**Human as a sensor** is one of the concepts of human involvement model defined in this dissertation. In this concept, the urban sensing system utilizes a sensory system of human (e.g., five senses) and psychological reactions (e.g., emotional status) as a new urban environmental sensor.

..... vi, 4, 5, 6, 8, 37, 38, 61, 101, 102, 132

**Human as an end-consumer** is one of the concepts of human involvement model defined in this dissertation. In this concept, the urban sensing system are unmanned and automated using Internet of Things (IoT) and sensor network etc.

..... 2, 132

---

**I**

---

**iBeacon** is a beacon concept utilizing Bluetooth Low Energy (BLE) technology developed by Apple Inc. It is Bluetooth advertising protocol for communicating between BLE devices, and can be used to identify and track a particular device, such as the enter/exit events which are fired when the user enters/exits the range of an iBeacon device.

..... 80

**Illuminance following (IF) characteristic** is a physical characteristic of illuminance found in this dissertation (see stColle:challenges). A smartphone embedded light sensor might not follow illuminance intensity changing correctly. It shows the tracking performance to illuminance changing.

..... 15, 17, 19, 21, 22, 28, 31

**Inertial measurement unit (IMU)** is an electronic device for measuring acceleration, angular rate, and sometimes magnetic field surrounding it, by using a combination of accelerometers and gyroscopes. It can be used for estimating the posture of the device and navigating without GPS. A major disadvantage of it is that it often suffers from accumulated error.

..... 60

**Information and Communication Technology (ICT)** is a general term of information technology and communication technology, and services, industry, facilities using them. This includes the Internet, wireless networks, cell phones, and other communication mediums.

..... 1, 8, 93, 98, 99, 102, 129

**Internet of Things (IoT)** is the network of various things which allows these connect, interact, and exchange data each other. The meaning of “things” is wide compared with the sensor network (e.g., devices, vehicles, home appliances, software, actuators).

..... 1, 2, 133, 135

**Inverse-square law of light** is a general physical characteristic of illuminance. The illuminance is attenuated with inversely proportional to the square of the distance from the light source ( $E_2 = \frac{1}{(d_2/d_1)^2} \cdot E_1$ ).  
 ..... 124

**Inverter-type fluorescent light** is one of the fluorescent lights which can turn on using *inverter circuit* instead of starter. Its illuminance is brighter than starter-type fluorescent light, and controllable.  
 ..... 28, 31, 33, 34, 135

**iOS** is a mobile operating system developed by Apple Inc. It works on iPhone, iPad, and iPod Touch, and is the second most popular mobile operating system globally after Android. The application working on iOS can be developed using Swift, an open source language.  
 – <https://www.apple.com/ios/>  
 ..... 77

---

**J**

---

**Japanese Industrial Standards (JIS)** is the standards used for industrial activities in Japan, which defined by Japanese industrial standards committee (JISC). Over 10 thousand standards are defined under JIS. – <http://www.jisc.go.jp/>  
 ..... 11, 12, 13, 14, 17, 23, 24, 25, 30

**Japanese Industrial Standards Committee (JISC)** is Japanese national standardization body who plays a central role in developing standards in Japan covering a wide range of products and technologies from robots to pictograms. It is also responsible for Japan’s growing contribution to setting international standards through its work with the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). – <http://www.jisc.go.jp/>  
 ..... vi, 11, 134

**Japan Security Systems Association (JSSA)** is a Japanese public interest incorporated association which conducts research and investigation of crime prevention equipment, crime prevention system, and information security system. – <https://www.ssaj.or.jp/>  
 ..... ix, 9, 11, 13, 14, 35

**JSON (JavaScript Object Notation)** is a lightweight data-interchange format consisting of two structures: a collection of name/value pairs, and an ordered list of values. – <https://www.json.org/>  
 ..... 74, 82

---

L

---

**Lambert's cosine law** is a general physical characteristic of illuminance. The intensity of illuminance observed from the light source is directly proportional to the cosine of the angle  $\theta$  between the direction of the incident light and the surface normal ( $E_\theta = E_0 \cdot \cos \theta$ ).  
..... 15, 18, 126, 136

**LED light** is an electric light that produce light using light-emitting diode (LED). It has high directivity and high illuminance compared with starter-type fluorescent light and inverter-type fluorescent light.  
..... 13, 28, 31, 33, 35

**Likert scale** is a method to measure opinions, perceptions, and behaviors. In general, it provides a choice of five to seven or even nine precoded question with the neutral point being neither agree nor disagree, and it can be answered in ordinal scales of agreement/disagreement.  
..... 44, 93, 95, 96

**Low-Level Descriptor (LLD)** is descriptive data which can be extracted from the signal itself. For instance, when using OpenSmile software, audio specific LLDs such as *Pitch*, *Loudness*, *Critical Band spectral*, and *Voice quality* can be extracted.  
..... 46, 49

---

M

---

**Mean Absolute Error (MAE)** is a measure of a difference between two continuous data, without considering their direction.  
..... 54, 57, 61, 102

**Ministry of Land, Infrastructure, Transport, and Tourism** is a ministry of the Japanese government. The ministry oversees four external agencies: the Japan Coast Guard, the Japan Tourism Agency, the Japan Meteorological Agency, and the Japan Transport Safety Board. – <http://www.mlit.go.jp/>  
..... 44

**Monetary incentive** is one of the actual reward to motivate users. Due to the strong force of “money,” it can be easily used for encouraging users’ behavior and contribution. However, there is a limitation in such monetary resources, it is difficult to operate sustainably.  
..... 66, 67, 73, 81, 90

**MQ Telemetry Transport (MQTT)** is a machine-to-machine (M2M) connectivity protocol which employs the publish/subscribe model. It uses the *topic* as a component of the message to determine which message goes to which MQTT client (subscriber). With a

rapid growth of IoT, it is focused renewed attention because it can provide many-to-many and real-time communication. – <http://mqtt.org/>

..... 78, 79, 135, 136

**MQTT Broker** plays the centric role of the publish/subscribe model: receiving all messages from clients, filtering the messages, deciding the subscribers for each message, and sending the message to these subscribers. It can deal with a huge amount of messages from concurrently connected MQTT clients.

..... 78, 136

**Publish** is the operation of the MQTT protocol, sending data to MQTT Broker. The MQTT client (called publisher) can send data consisting of a *topic* and content body to MQTT Broker without specifying the destination.

..... 79, 135

**Subscribe** is the operation of the MQTT protocol, receiving data from MQTT Broker. By registering *topic* want to subscribe in advance, the MQTT client (called subscriber) can receive data without specifying the publisher.

..... 79, 135

---

## N

---

**Non-monetary incentive** is one of the virtual reward for motivating users to contribute to tasks. It gives “experience” instead of monetary rewards to users. As the “experience”, there are many options can be used, e.g., points in the app, ranking, visualization, data-sharing or communication with other users. The gamification mechanism is categorized as this incentive.

..... 66, 67, 81, 82, 84, 90, 91, 131

---

## O

---

**Obliquely incident light (OIL) characteristic** is a physical characteristic of illuminance found in this dissertation (see stColle:challenges). In general, when using illuminometer, it follows Lambert’s cosine law. However, when using a smartphone embedded light sensor, it shows a different characteristic, due to its shape of the sensor surface.

..... viii, 15, 17, 18, 19, 20, 21, 24, 28, 30, 31, 125, 126

**Open data** is data that can be freely used, reused, and redistributed and built-on by anyone, anywhere, for any purpose. Many of the data is published under some kind of licenses such as GPL (GNU General Public License), CC (Creative Commons), and ODbL (Open Data Commons Open Database License). – <https://okfn.org/opendata/>

..... 10, 35

**Opportunistic sensing** is a crowdsourcing-based sensing method which the system implicitly performs sensing tasks after getting user’s agreement.

..... 4, 5

---

**P**

---

**Participatory sensing** is a crowdsourcing-based sensing method which the user intentionally performs sensing tasks based on the request from the system.

..... 3, 4, 5, 62, 63, 65, 67, 68, 70, 71, 72, 74, 77, 80, 83, 84, 86, 88, 89, 91, 98, 99, 102

**Point of Interest (POI)** is a specific point location on a map instead of linear objects such as roads. It is not necessarily an interesting place. For instance, public facility, restaurant, historical statue, car parks, and tourist attractions etc. can be assigned as POI.

..... 88

***p*-value** is the probability value to be used to determine significance on null hypothesis significance testing. This value will be larger when the null hypothesis is true. In general, If this value becomes lower than 0.01, 0.05, or 0.1 ( $p < 0.01$ ,  $p < 0.05$ , or  $p < 0.1$ ), it can be determined that there is a strong, moderately, or weak significant difference.

..... 58

---

**Q**

---

**QR (Quick Response) code** is a matrix barcode (two-dimensional barcode) invented by DENSO Corporation. It consists of black squares arranged in a square grid on a white background, which can be read by an imaging device such as a camera.

..... 74, 78, 96

---

**R**

---

**RAMAS (The Russian Acted Multimodal Affective Set) database** is a play-acted multimodal database which consists of 150 sessions of video, audio, motion capture and psychophysiology (PPG, GSR) recordings of improvised affective dyadic interactions. The seven emotions (joy, anger, sadness, disgust, fear, surprise, and neutral) are played by 10 semi-professional actors (five men and five women, 18–28 years old, native Russian speakers). – <https://neurodatalab.com/ramas.html>

..... 49, 50

**RDB (Relational Database)** RDB is a set of data stored as tables, records, and columns. Even the complex datasets can be organized clearly by defining relationships between each dataset. Generally, it uses SQL (Structured Query Language) to access the datasets.

..... 74, 82

**RECOLA (Remote COLlaborative and Affective interactions) database** is a multi-modal database which consists of 9.5 hours of audio, visual, and physiological (electrocardiogram, and electrodermal activity) recordings of online dyadic interactions between 46 French-speaking participants. – <https://diuf.unifr.ch/main/diva/recola/>  
 ..... 49, 50

**Rectified Linear Unit (ReLU)** is one of the activation function of a node (neuron) in neural networks, which defines the output of that node for inputs. ReLU uses *Ramp function* ( $\varphi(x) = \max(0, x)$ ).  
 ..... 50

**Recurrent Neural Networks with Long Short-Term Memory (RNN-LSTM)** is the recurrent neural network (RNN) which has the long short-term memory (LSTM). The RNN is an artificial neural network which has the time-sequential and backward connection between hidden layers. It can be used effectively for time series analysis, natural language processing, and speech recognition. The LSTM network is designed to avoid long-term dependencies problem of RNN, and capable of learning with remembering information for long periods.  
 ..... 46, 50

**Russell’s circumplex space model** is the representation of emotion with a two-dimensional circular space, containing arousal and valence dimensions. The vertical and horizontal axis show the arousal and valence respectively, and the center of the circle represents a neutral valence and a medium level of arousal.  
 ..... 44

**Arousal** represents the intensity of psychological state. For instance, “angry” or “excited” has high intensity and high arousal state, also, “bored” or “calm” has low intensity and low arousal state. Each emotion has a different valence state.  
 ..... 51, 138

**Valence** represents the positive-negative level of psychological state. For instance, “delighted” or “relaxed” has a positive (high) valence, “frustrated” or “depressed” has a negative (low) valence. Each emotion has a different arousal state.  
 ..... 51, 138

---

S

---

**SEMAINE (Sustained Emotionally coloured MACHine-human Interaction using Non-verbal Expression) database** is an audiovisual database constructed for a project whose aim was to build a system that could engage a person in a sustained conversation with a SAL (Sensitive Artificial Listener) agent. It consists of 24 recordings

in English from 20 speakers aged between 20 and 58 years (7.2 hours in total). – <a href="https://semaine-db.eu/">https://semaine-db.eu/</a>	49, 50
<b>Sensor network</b> is a sensing method for monitoring and recording the physical conditions of the target environment using a group of spatially distributed and dedicated sensors.	2, 3, 132, 133
<b>Similarity of illuminance transition (SoIT)</b> is a metric to evaluate the accuracy of illuminance transition fitting in this dissertation (see Section 2.4.1). It is the sum of absolute error of illuminance value for each measurement data between the correction standard-/target.	30, 31
<b>Social Networking Service (SNS)</b> is an online platform which general people can communicate with each other for building social relations with other people. It supports finding the person who has similar interests, career, activities, backgrounds, and real-life connections.	3, 39, 132, 139
<b>Social sensing</b> is a sensing method for collecting data from humans or devices on their behalf. Due to the widespread of mobile internet, people can share real-time experiences through the social networking service (SNS) etc. It is an approach extracting useful information from such unspecified large number of data sources posted by a general people.	3
<b>Spoken-dialogue system</b> is a method for convenient interaction between user and system with voice, such as performing information retrieval task, using services, and having a conversation.	38
<b>SQLite</b> is a relational database engine using SQL (Structured Query Language). – <a href="https://www.sqlite.org/">https://www.sqlite.org/</a>	82
<b>Stamp-rally</b> is a motivating approach used to entice users to go around to certain places at events etc. By giving a motivation to complete to collect physical/digital <i>stamps</i> which are placed in whole events area, it can induce the user’s behavior implicitly.	91, 95
<b>Starter-type fluorescent light</b> is one of the fluorescent lights which needs <i>starter</i> to allow current to flow through the filaments at the ends of the tube.	28, 33, 134, 135

**Student's *t*-test** is statistical testing method which is used to determine whether there is a significant difference between the means of two groups which are related in certain features.  
 ..... 56, 57, 58

---

**U**

---

**Ubiquitous computing** is a concept of computer science which means computing is available anywhere and anytime in our daily life. *Ubiquitous* is derived from the Latin word “*ubique*” which means omnipresent. This paradigm is also described as pervasive computing.  
 ..... 1

**Unweighted Average Recall (UAR)** is a measure to evaluate with unbalanced classes. The accuracy per class divided by the number of classes without considerations of instances per class.  
 ..... 52, 54, 56, 57, 61, 102

---

**W**

---

**Weighted Moving Average (WMA)** WMA is a method for smoothing series data (generally, time-series data). It calculates a weighted average of the last *n* values.  
 ..... 19, 22

---

**X**

---

**XML (Extensible Markup Language)** is a markup language that defines a set of rules for encoding documents in both human- and machine-readable text format. – <https://www.w3.org/XML/>  
 ..... 82

# Publication List

## Peer Review Journal Paper

1. Yuki Matsuda, Dmitrii Fedotov, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, Wolfgang Minker: “EmoTour: Estimating Emotion and Satisfaction of Users Based on Behavioral Cues and Audiovisual Data,” *Sensors* 2018, Vol.18, No.11, Article ID 3978, 19 pages, 2018, doi: 10.3390/s18113978.
2. Yuki Matsuda, Ismail Arai, Yutaka Arakawa, Keiichi Yasumoto: “A Correction Method for Individual Differences among Smartphones’ Light Sensor to Measure Streetlamps’ Illuminance for Sidewalks at Night,” *IPSJ Journal*, Vol.57, No.2, pp.520–531, 2016. (in Japanese)
3. Yutaka Arakawa, Yuki Matsuda: “Gamification mechanism for enhancing a participatory urban sensing: survey and practical results,” *Journal of Information Processing*, Vol.57, No.1, pp.31–38, 2016, doi: 10.2197/ipsjjip.24.31.

## International Conference

1. Dmitrii Fedotov, Yuki Matsuda, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, Wolfgang Minker: “Towards Real-Time Contextual Touristic Emotion and Satisfaction Estimation with Wearable Devices,” 2019 IEEE International Conference on Pervasive Computing and Communications Workshops PerCom Workshops ’19), pp.358–360, 2019.
2. Dmitrii Fedotov, Yuki Matsuda, Wolfgang Minker: “From Smart to Personal Environment: Integrating Emotion Recognition into Smart Houses,”

The 1st International Workshop on Pervasive Computing and Spoken Dialogue Systems Technology (PerDial '19), in conjunction with 2019 IEEE International Conference on Pervasive Computing and Communications Workshops PerCom Workshops '19), pp.943–948, 2019.

3. Kenta Taki, Yuki Matsuda, Yutaka Arakawa, Keiichi Yasumoto: “Design and Implementation of Notification Information Survey System and Survey Results Toward Use-side Adaptive Notification Management,” The first International Workshop on Pervasive Persuasive System for Behavior Change (PerPersuasion '19), in conjunction with 2019 IEEE International Conference on Pervasive Computing and Communications Workshops PerCom Workshops '19), pp.808–813, 2019.
4. Yuki Matsuda, Dmitrii Fedotov, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, Wolfgang Minker: “EmoTour: Multimodal Emotion Recognition using Physiological and Audio-Visual Features,” Ubiquitous Emotion Recognition with Multimodal Mobile Interfaces (UERMMI), in Adjunct Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp/ISWC '18 Adjunct), pp.946–951, 2018, doi: 10.1145/3267305.3267687.
5. Dmitrii Fedotov, Yuki Matsuda, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, Wolfgang Minker: “Towards Estimating Emotions and Satisfaction Level of Tourist based on Eye Gaze and Head Movement,” Proceedings of the 2018 IEEE International Conference on Smart Computing (SMARTCOMP '18), pp.399–404, 2018, doi: 10.1109/SMARTCOMP.2018.00036.
6. Shogo Kawanaka, Yuki Matsuda, Hirohiko Suwa, Manato Fujimoto, Yutaka Arakawa, Keiichi Yasumoto: “A gamified participatory sensing for tourism: the effect to a sightseeing,” The Third International Workshop on Smart Sensing Systems (IWSSS '18), pp.196–205, 2018, doi: 10.3233/978-1-61499-874-7-196.

7. Yuki Matsuda, Dmitrii Fedotov, Yuta Takahashi, Yutaka Arakawa, Keiichi Yasumoto, Wolfgang Minker: “Estimating User Satisfaction Impact in Cities using Physical Reaction Sensing and Multimodal Dialogue System,” The Ninth International Workshop on Spoken Dialog System Technology (IWSDS '18), 6 pages, 2018.
8. Masato Hidaka, Yuki Matsuda, Shogo Kawanaka, Yugo Nakamura, Manato Fujimoto, Yutaka Arakawa, Keiichi Yasumoto: “A System for Collecting and Curating Sightseeing Information toward Satisfactory Tour Plan Creation,” The Second International Workshop on Smart Sensing Systems (IWSSS '17), 2017.
9. Yuki Matsuda, Takashi Hasegawa, Ismail Arai, Yutaka Arakawa, Keiichi Yasumoto: “WaistonBelt 2: A Belt-type Wearable Device for Monitoring Abdominal Circumference, Posture and Activity,” The 9th International Conference on Mobile Computing and Ubiquitous Networking (ICMU '16), pp.1–6, 2016,  
doi: 10.1109/ICMU.2016.7742089.
10. Yuki Matsuda, Yutaka Arakawa, Keiichi Yasumoto, “Design and Evaluation of Participatory Mobile Sensing Platform for Diverse Sensing and Gamification Scenarios,” ASSET Poster on the 14th ACM International Conference on Mobile Systems, Applications, and Services (MobiSys '16), p.57, 2016.
11. Yuki Matsuda, Takashi Hasegawa, Keiichiro Iwanami, Naoya Saito, Takuya Ishioka, Ismail Arai, Yutaka Arakawa, Keiichi Yasumoto: “WaistonBelt: A Belt for Monitoring Your Real Abdominal Circumference Forever,” Demonstration on the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15), pp.329–332, 2015,  
doi: 10.1145/2800835.2800869.

## Domestic Conference in Japanese

1. 高城賢大, 松田裕貴, 諏訪博彦, 藤本まなと, 荒川豊, 安本慶一: “端末側での適応的通知タイミング制御の実現に向けた通知情報調査システムの構築と

調査結果,” 情報処理学会研究報告 (2018-UBI-60) , 2018.

2. 河中祥吾, 松田裕貴, 諏訪博彦, 藤本まなと, 荒川豊, 安本慶一: “観光客参加型センシングによる観光情報収集におけるゲーミフィケーションの有効性調査,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO 2018) シンポジウム論文集, 2018.
3. 高橋雄太, 松田裕貴, Dmitrii Fedotov, 荒川豊, Wolfgang Minker, 安本慶一: “観光中の内的状態推定に向けた観光客の無意識的しぐさの分析,” 電子情報通信学会技術研究報告, ヒューマンプローブ研究会 (HPB) , 2018.
4. 松田裕貴, 河中祥吾, 諏訪博彦, 荒川豊, 安本慶一: “ユーザ参加型センシングの割り込みに対する応答性調査 ～時空間データとタスク難易度およびユーザ属性による考察～,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO 2017) シンポジウム論文集, 2017.
5. 河中祥吾, 松田裕貴, 藤本まなと, 荒川豊, 安本慶一: “スマートフォンで撮影した近接群衆画像からの混雑度推定手法に関する一検討,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO 2017) シンポジウム論文集, 2017.
6. 日高真人, 松田裕貴, 河中祥吾, 中村優吾, 藤本まなと, 荒川豊, 安本慶一: “実時間観光コンテンツ提供に向けた観光情報収集・キュレーションシステムの提案,” 第68回 高度交通システムとスマートコミュニティ研究会 (ITS) , 2017.
7. 松田裕貴, Akpa Akpro Elder Hippocrate, Konan N'djabli Cedric Ange, 中村優吾, 前田直樹, 千住琴音, 荒川豊: “位置情報サービスにおける個人特化型ゲーミフィケーション ～スタンプラリーイベントを通じた「慣れ」「飽き」の調査～,” 第3回行動変容と社会システム研究会 (BCSS) , 2017.
8. 松田裕貴, 荒川豊, 安本慶一: “多様なユースケースに対応可能なユーザ参加型モバイルセンシング基盤の実装と評価,” 情報処理学会, マルチメディア, 分散, 協調とモバイル (DICOMO 2016) シンポジウム論文集, 2016.
9. 松田 裕貴, 荒川 豊, 安本 慶一: “DyPRoS: 地域情報発信のための動的フォトロゲイニングシステム,” 電子情報通信学会技術研究報告, HCG シンポジウム 2015, 2015.

10. 中村 優吾, 松田 裕貴, 荒川 周造, 金平 卓也, 安本 慶一：“GAIFoT：情報流を活用した地域分散型アンビエントインターフェース,” ポスター発表, 第23回マルチメディア通信と分散処理ワークショップ (DPSWS 2015) 論文集, 2015.
11. 中村 優吾, 松田 裕貴, 荒川 周造, 金平 卓也, 安本 慶一：“屋内空間の状況推定に基づく移動支援システムの提案,” ポスター発表, 電子情報通信学会技術研究報告, コンピュータシステム研究会 (CPSY) , 2015.
12. 松田裕貴, 新井イスマイル, 荒川豊, 安本慶一：“街灯これくしょん：夜道の明るさをセンシングする参加型アプリケーション,” 電子情報通信学会技術研究報告, ヒューマンプローブ研究会 (HPB) , 2015.