Estimation of Thermal Sensation Based on Machine Learning via Physiological Sensing

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2022

Hiroki YOSHIKAWA

List of Publications

Related Journal Articles

(1) Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "TSVNet: Combining Time-Series and Opportunistic Sensing by Transfer Learning for Dynamic Thermal Sensation Estimation," *IEEE Access*, vol. 9, pp. 102835–102846, 2021.

(2) Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "ThermalWrist: Smartphone Thermal Camera Correction Using a Wristband Sensor," *Sensors*, vol. 19, no. 18, p. 3826, 2019.

Related Conference Papers

(1) Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "Data Balancing for Thermal Comfort Datasets Using Conditional Wasserstein GAN with a Weighted Loss Function," in *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 264–267, November 2021.

(2) Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "Time-Series Physiological Data Balancing for Regression," in *Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Computer Applications*, pp. 393 – 398, June 2021.

(3) Hiroki Yoshikawa, Akira Uchiyama, Yuki Nishikawa, Teruo Higashino, "Combining a thermal camera and a wristband sensor for thermal comfort estimation," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 238–241, September 2019.

(4) Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "Dynamic Offset Correction for Smartphone Thermal Cameras Using a Wristband Sensor," in *Proceedings of the 2019 IEEE International Conference on Pervasive Computing and Communications Workshops*, pp. 165–170, March 2019.

Other Conference Papers

(1) Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "Reliability Estimation of Heart Rate Measurement Using Wrist-Worn Devices," in *Proceedings of the 13th International Conference on Mobile Computing and Ubiquitous Networking*, pp. 1–6, November 2021.

(2) Fumika Beppu, Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, Keisuke Hamada, Eiji Hirakawa, "Body Part Detection from Neonatal Thermal Images Using Deep Learning," in *Proceedings of the 18th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pp. 1–13, November 2021.

(3) Masayuki Hayashi, Hiroki Yoshikawa, Akira Uchiyama, Teruo Higashino, "Preliminary Investigation on Band Tightness Estimation of Wrist-Worn Devices Using Inertial Sensors," in *Proceedings of the 10th EAI International Conference on Wireless Mobile Communication and Healthcare*, pp. 256–266, November 2019.

Abstract

A comfortable air-conditioned environment increases labor productivity and learning efficiency. In indoor environments used by many people, such as office buildings and schools, it is necessary to realize an appropriate air-conditioned environment that considers users' thermal sensation. However, it is difficult to always provide the air conditioning control required by the users in these facilities. This is because the ideal environment changes dynamically due to spatial heterogeneity of the environment, outdoor weather changes, and human density changes. In addition, it is difficult to realize an air-conditioned environment that considers all the people's preferences in the same space by using the existing air conditioning control. This is because thermal sensation differs among individuals due to the effects of clothing and metabolism, even in the same air conditioned environment.

Many researchers proposed systems using machine learning to estimate the dynamically changing thermal sensation of individuals using physiological sensing devices. In addition to environmental factors, i.e., temperature and humidity, these devices can be used to estimate thermal sensation with data collected from a thermal camera and a wearable sensor. The thermal sensation is useful to determine comfortable settings for air conditioning systems. However, some problems are not addressed in the flow of the construction of the machine learning model to estimate thermal sensation.

A thermal camera is widely used to measure human body temperature for various healthcare applications, including thermal sensation estimation. Smartphone thermal cameras are in the market as mobile thermal sensing devices. However, thermal images captured by smartphone thermal cameras are not accurate in monitoring human body temperature due to the small body that is vulnerable to temperature change.

On the other hand, the development of sensing technologies on the high-end device has enabled the estimation of human thermal sensation, which helps efficient heating, ventilation, and air conditioning control. However, in most existing literature, the personal thermal sensation is estimated using only opportunistic data as features, and evaluations are conducted in static environments. In real situations, we are exposed to dynamic environments, owing to human mobility and changes in airflow. Therefore, it is necessary to estimate personal thermal sensation in dynamic surrounding environments.

In addition, even using an appropriate estimation model, dataset imbalance causes biased estimation, which is harmful for the estimation of rare cases. Thermal sensation datasets are usually imbalanced because hot/cold environments rarely appear in air-conditioned environments. Therefore, many

researchers have applied data augmentation for rare samples to balance thermal sensation datasets, called data balancing. In physiological sensing, estimations of human states such as thermal sensation using machine learning often regard the target problems as classification problems for simplicity. However, to achieve estimations of human states with finer granularity, regression is more appropriate than classification because regression estimates numerical values. Because regression problems output continuous numerical values, data balancing algorithms for classification problems are not suitable for regression problems. Some algorithms are proposed for data balancing in regression problems. However, they do not consider time-series feature values commonly used in psychological and physiological state estimation.

Our research goal is to build a machine learning model to estimate thermal sensation. Building the model requires three key steps: the physiological data collection, the selection and construction of the machine learning model, and the training schemes to address the imbalanced dataset.

Firstly, we propose ThermalWrist, a dynamic offset correction method for thermal images captured by smartphone thermal cameras. We fully utilize the characteristic which is specific to thermal cameras: the relative temperatures in a single thermal image are highly reliable, although the absolute temperatures fluctuate frequently. ThermalWrist combines thermal images with a reliable absolute temperature obtained by a wristband sensor based on the above characteristic to correct the offset error. The evaluation result in an indoor air-conditioned environment shows that the mean absolute error and the standard deviation of face temperature measurement error decrease by 49.4% and 64.9%, respectively. In addition, Pearson's correlation coefficient increases by 112%, highlighting the effectiveness of ThermalWrist. We also investigate the limitation with respect to the ambient temperature where ThermalWrist works effectively. The result shows ThermalWrist works well in the normal office environment, which is 22.91°C and above.

Secondly, we propose TSVNet, a deep learning-based method reflecting time-series changes to address the dynamic environment. This method combines opportunistic features and time-series features in the deep learning framework by transfer learning. We collected data for a total of 123 days, which included the dynamic environment data from 21 subjects, for the evaluation. The results indicate that our method improves F1-score by 5.8% compared with a baseline method. We also design a data balancing method for regression problems on imbalanced datasets, including time-series data. In addition, the result of the lookback time evaluation shows that the use of physiological information in the previous 10 minutes improves the performance of the method.

Thirdly, we propose a SMOTE-based method to alleviate the distribution bias by data augmentation in the regression problem using a time-series physiological data dataset. To consider the temporal dependency of time-series data, we extend a distance function to measure the distance between two samples in the dataset. We use Dynamic Time Warping (DTW) distance to define the distance between time-series samples. Our method interpolates synthetic time-series using the DTW distance between rare samples. The effectiveness of the proposed method was confirmed for datasets of thermal

sensation and core body temperature collected in uncontrolled environments. The results show that our method improves the performance of regression models for minor cases with a bit of decline in the mean average error.

We propose a framework to build a machine learning model to estimate thermal sensation from the data collection to train the model through these contributions. The complete system will appropriately estimate rare cases, i.e., hot and cold, which possibly include extra energy use. This dissertation has established how we utilize our physiological data to build a machine learning model that estimates our thermal sensation.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

In recent years, as mentioned in the Sustainable Development Goals (SDGs) [1], which are an urgent call for action by all countries in a global partnership, energy consumption all over the world continues to increase. Since the air-conditioning application accounts for a high percentage of all energy consumption [2], the reduction of extra energy consumption related to air-conditioning control, such as overcooling, is expected to have a significant effect on reducing the energy consumption of the building. In addition, it has been found that comfortable air-conditioned environments increase labor productivity [3] and learning efficiency [4]. In indoor environments used by many people, such as office buildings and schools, it is necessary to create appropriate air-conditioned environments in terms of users' thermal sensation. However, it is not easy to always control the air conditioning system in these buildings according to users' demands. This is because the ideal environments change dynamically due to spatial heterogeneity, outdoor weather changes, and human density changes. It is also challenging to realize air-conditioning environments that consider all the people's preferences in the same space by using the existing air-conditioning control system. This is because thermal sensation varies from person to person due to the effects of clothing and metabolism, even in the same air-conditioning environment. In recent years, there has been much research on estimating the dynamically changing thermal sensation of individuals using sensing devices to tackle such challenges.

Many researchers proposed systems using machine learning to estimate the dynamically changing thermal sensation of individuals using physiological sensing devices. In addition to environmental factors, i.e., temperature and humidity, these devices can be used to estimate thermal sensation with data collected from a thermal camera and a wearable sensor. The thermal sensation is useful to determine comfortable settings for air conditioning systems. However, some problems are not addressed in the flow of the construction of the machine learning model to estimate thermal sensation.

Data collection is another challenge in thermal sensation estimation. Specifically, a thermal camera is a sensing device that helps estimate thermal sensation because it provides facial temperature data, sensitively reflecting human thermal sensation. In addition, psychological states, for example, cognitive load [5], thermal sensation [6, 7], stress [8] and emotion [9], are estimated by monitoring the

skin temperature. Although some wearable devices with skin temperature sensors are available (e.g., E4 wristband [10], and Embrace 2 [11]), their applications are still limited due to the limitation of single point measurement. For this reason, thermal cameras are widely used to monitor skin temperature because they can measure temperature distributions quickly without physical contact. Recently, thermal cameras have become easy to use because low-cost and minimized models are available on the market. For example, FLIR ONE [12] is a low-cost thermal camera that is attached to a smartphone. We can measure our skin temperature anytime, anywhere thanks to such smartphone thermal cameras. However, the accuracy of the low-cost thermal camera is insufficient to monitor the skin temperature compared with the high-end models. Thermal cameras are categorized into two types: those with cooled infrared detectors and those with uncooled infrared detectors [13]. The performance of the cooled detectors is much higher than the other, although they are bulky and expensive due to the cooling apparatus. Therefore, the smartphone thermal cameras are with uncooled infrared detectors. The uncooled infrared detector element converts its temperature rise to electric signals. The object's temperature is principally calculated from the signals and their emissivity. The measured temperature greatly fluctuates by the parameters configured by the user, the rising temperature of the camera's body, the efficiency of the element, and the packaging method [13]. By these effects, the measurement error of the smartphone thermal camera is larger than the high-end one. For example, the error range is $\pm 3^{\circ}$C or $\pm 5\%$, which is larger than the range of human skin temperature changes in daily life. This is clearly not enough for various healthcare applications. The dynamic correction method for smartphone thermal cameras is needed for human thermal sensing to address this challenge.

Some smart ACs reduce energy consumption by controlling the airflow based on the user's position estimated by an infrared sensor. However, appropriate control is non-trivial because thermal sensation inherently contains subjective feelings. Fanger proposed an index called PMV (Predicted Mean Vote) that represents human thermal sensation in 7 levels from -3 to +3 (i.e., cold, slightly cool, neutral, slightly warm, warm, hot) [14]. PMV is widely used in recent research as a 7-point ASHRAE (American Society of Heating, Refrigerating and Ari-Conditioning Engineers) scale, an international air conditioning association. PMV calculates thermal sensation from environmental factors, such as temperature, humidity, wind speed, radiation temperature, and personal factors such as metabolism and clothing. However, it has been reported that the thermal sensation calculated using PMV is not necessarily equal to the thermal sensation actually felt by humans because PMV does not consider individual differences appropriately and assumes the environment is in a steady state. In order to consider such subjective factors, TSV (Thermal Sensation Vote) [15] is often used as a thermal sensation index. To estimate TSV, Ranjan et al. proposed a contact-less estimation method based on machine learning from air temperature and skin temperature [6]. However, a thermal camera requires capturing the user's face, which is not always possible due to the camera angle, user's body/face direction, etc. On the other hand, Barrios et al. used heart rates measured by a chest heart rate monitor to estimate the thermal sensation [16]. Although it can continuously monitor users' physiological data, wearable

sensors such as a wristband sensor are preferable in daily use.

Many studies have shown the effectiveness of machine learning in estimating psychological or physiological states using physiological data as input. However, it is ethically and physically difficult to collect a large amount of data without bias in uncontrolled environments. Specifically, the amount of data in rare cases is especially small compared to common data. Therefore, the distribution bias may cause overfitting in machine learning. For example, in the thermal sensation dataset, hot/cold is rarely reported compared to the response labeled neutral. When we use such a dataset to train the machine learning model, the estimator tends to output the neutral label. To address this problem, many data balancing methods are proposed. Data balancing is a way of pre-processing the data to increase its size, diversity, and robustness. Synthetic Minority Oversampling Technique (SMOTE) [17] is one of the methods to balance the dataset. SMOTE augments the data in minor classes by interpolation for classification problems. To estimate the physiological or psychological state such as the thermal sensation, regression is more appropriate than classification because regression estimates such state as continuous numerical values. As a data balancing for regression problems, SMOTER [18] were proposed based on SMOTE. However, it does not assume that time-series feature as an input, which is commonly used in psychological or physiological state estimation. Therefore, to balance the regression dataset, including time-series data, we need to design the framework to augment time-series data, which helps to build an estimator that executes balanced estimation.

In this dissertation, three primary contributions will be made. We tackle three key challenges to build the machine learning model for TSV estimation: accurate physiological data collection, selection and construction of the machine learning model, and the training schemes to address the imbalanced dataset.

Firstly, to overcome the fluctuation of the measurement by smartphone thermal cameras, we propose an offset correction method for thermal images captured by smartphone thermal cameras. We fully utilize the key feature of the thermal cameras: the measurement fluctuation is mainly caused by the offset, which is common in all the pixels in a single thermal image. In other words, we can measure the difference of temperature correctly between any pair of pixels in the same thermal image, even using the smartphone thermal camera. Our method combines thermal images with a reliable absolute temperature obtained by a wristband sensor based on the above feature. First, we obtain a thermal image including a reference point (a wrist or a palm) of which absolute temperature is measured by a wearable device (a wristband sensor). Second, we estimate the offset at the reference point by comparing the temperature measured by the thermal camera and the wristband sensor. Finally, our method corrects temperature in the thermal image by adding the offset to all the pixels. Since the wristband sensor covers the measurement point, the thermal camera cannot directly capture the temperature of the same point. Therefore, we define the reference point as the point that is highly correlated with the point measured by the wristband sensor. This method uses a palm or a wrist as the reference point and compares the performance through the real experiment. We compare the two reference points and

discuss their features through the real data with 876 samples collected from eight subjects.

Secondly, we propose TSVNet, which combines opportunistic data from a thermal camera with time-series data from a wristband sensor for thermal sensation estimation. TSVNet estimates users' TSV based on their physiological data and face temperature distribution by machine learning. TSVNet is a deep learning-based method reflecting time-series changes to address dynamic environments. The basic idea is as follows. First, we train the partial estimator using time-series data only. Second, to combine the knowledge extracted from the time-series data with the opportunistic data, TSVNet transfers the partial network of the estimator for the time-series data to the input of the latter combined TSV estimator. Finally, the transferred network is combined with the opportunistic data as the input to the latter combined TSV estimator. We collected data for a total of 123 days, which included data collected under dynamic environments from 21 subjects for the evaluation. Because our experiment was conducted in realistic environments, most of the TSVs labeled by participants are 0 (Neutral). This makes the estimation more challenging compared with the existing state-of-the-art studies, which divide the occupant's thermal state into comfort and discomfort [19], [20], [21]. Because such an imbalanced dataset is used, we also propose a data balancing method for the time-series data in the regression problem to prevent biased learning. The results indicate that our method improves F1-score by 5.8% compared with a baseline method. We also design a data balancing method for regression problems on imbalanced datasets, including time-series data. In addition, the result of the lookback time evaluation shows that the use of physiological information in the previous 10 minutes improves the performance of the method.

Thirdly, we propose a data balancing method for regression with time-series feature values based on SMOTER. We extend a distance function to consider the temporal dependency of time-series data. To define the distance between time-series samples, we use Dynamic Time Warping (DTW) distance as used in TS_SMOTE [22]. TS_SMOTE is designed to extend SMOTE to time-series data. Our method interpolates synthetic time-series using the DTW distance. We apply our method to two imbalanced datasets with time-series feature values for evaluation. The first dataset is the thermal sensation dataset, which consists of time-series physiological data measured by a wristband sensor as a feature value and thermal sensation vote (TSV) as a response value. The second dataset is the core body temperature dataset. A tympanic temperature sensor measures the core body temperature during exercise. Feature values are measured by a wristband sensor, chest strap sensor, and environmental sensor. The result shows that our method improves the performance of regression models for rare cases with a little decline in the mean average error.

Our study overcomes challenges in order to build a machine learning model to estimate human thermal sensation by physiological sensing. The key contributions in this thesis are as follows.

- We propose an offset correction method for thermal images measured by a smartphone thermal camera using a reference point combined with the specific features of thermal cameras.

- We construct the estimator that combines the physiological information acquired as time-series data and the opportunistically measured data using transfer learning.
- We propose a data balancing method for regression with time-series feature values considering the temporal dependency of time-series data.

The rest of this dissertation is organized as follows. Chapter 2 reviews related work on thermal sensing, thermal sensation estimation, and data balancing method. Chapter 3 explains the design and performance of our correction method for smartphone thermal cameras. Chapter 4 proposes the thermal sensation estimator, which combines opportunistic and time-series data. Chapter 5 describes the novel method to balance the regression dataset, including time-series data. Finally, Chapter 6 summarizes and concludes this dissertation.

Chapter 2

Related Work

2.1 Related Work on Thermal Camera Sensing

2.1.1 Applications Using Skin Temperature

Many research has revealed the relation between human mental states and skin temperature. Choi et al. [23] investigate the possibility of the use of human body skin temperature to assess thermal sensation using the temperature sensor on the skin. Chernyshov et al. [24] present the system for tracking positive cognitive and emotional states by using temperature sensors on eyeglasses. Genno et al. [8] use facial skin temperature to evaluate stress and fatigue. They revealed the fatigue is the load of the stress by assuming the estimation formula. Understanding the stress of humans is a crucial issue in our society. Recently, Japanese companies have been obligated to conduct the stress check test for the employees by law [25]. As used in the work mentioned above, thermal cameras are one of the key technologies to record stress levels noninvasively and automatically. In addition, smartphone thermal cameras extend the advantages of thermal cameras in terms of ubiquitous applications.

2.1.2 Applications Using High-end Thermal Camera

Thermal cameras have been recently used for the estimation of human thermal comfort. Burzo et al. [7] divide the thermal comfort into three levels: "hot discomfort," "comfort," and "cold discomfort" and combine other biosensors with a thermal camera to estimate the thermal comfort. Ranjan et al. [6] estimate the thermal sense using thermal images and propose a method to reduce energy consumption in buildings. Stress or cognitive load estimation methods are shown by many researches [5, 26, 27, 28, 8, 29, 30]. Abdelrahman et al. [5] present an unobtrusive indicator of users' cognitive load based on thermal images by monitoring forehead and nose temperature. The other emotions are estimated by monitoring the skin temperature [31]. Pavlidis et al. [32] detect lies based on bloodstream increase estimated by thermal images. López et al. [33] and Basu et al. [9] propose methods to estimate human emotion by facial temperature distribution. Thermal cameras are also useful to measure such psychological states since they do not disturb user behavior.

Table 2.1: Thermal cameras used in existing work

Author	Thermal camera	Accuracy
Abdelrahman et al. [5]	Optris PI160	±2 °C or ±2%
Or et al. [27]	MikroScan 7200V	±2 °C or ±2%
Kang, et al. [28]	MikroScan 7200V	±2 °C or ±2%
Shastri et al. [29]	FLIR SC6000	±2 °C or ±2%
Ranjan et al. [6]	FLIR A655sc	±2 °C or ±2%
Burzo et al. [7]	FLIR A40	±2 °C or ±2%
Basu et al. [9]	InfReC R300	±1 °C or ±1%
Gane et al. [34]	FLIR SC640	±2 °C or ±2%
Ebisch et al. [35]	FLIR SC3000	±1 °C or ±1%
Manini et al. [36]	FLIR SC660	±1 °C or ±1%
Hahn et al. [37]	testo 881	±2 °C or ±2%

However, most of the existing work, shown in Table 2.1, use a high-end thermal camera that has high accuracy, such as ±2 °C or ±2%, or higher. In reality, we can not always use such high-end thermal cameras, causing the problem of frequent fluctuations of measurement by low-cost (i.e., smartphone) thermal cameras.

2.1.3 Applications Using Smartphone Thermal Camera

Some applications using smartphone thermal cameras have been developed since their release. Nurmi et al.[38] developed a low-cost solution for search and rescue operations using smartphone thermal cameras. Mauriello et al.[39] evaluated energy-efficiency issues in building environments. However, they use smartphone thermal cameras for measuring objects whose temperature range is much wider than human skin temperature. In addition, there are some applications for medical supports. For example, Jaspers et al.[40] use the smartphone thermal camera for the burn wound assessment. However, they measure temperature differences between the burn wound and healthy skin. As far as we know, there are no applications to measure human skin temperature using only smartphone thermal cameras. This is because smartphone thermal cameras are suitable for the measurement of relative temperature rather than absolute temperature.

2.1.4 Calibration of Thermal Cameras

Thermal cameras need to calibrate [41] to measure accurate temperatures. The purpose of calibration is to determine the accurate quantitative relations between camera output and incident radiation. Malmivirta et al.[42] mitigate the smartphone thermal camera's error using a deep learning-based calibration technique. They achieve smartphone thermal camera calibration without any additional device. However, the error larger than 1 °C still remains. Our target scenarios are measuring human skin temperature change over time, which requires higher accuracy. For the calibration procedure [13], different temperature black bodies are used, whose emissivity is close to unity and radiometric quan-

tities and spectral quantities are well defined. Therefore, the calibration method gives us relations of proper output and the black body's temperature. Instead of the blackbody, shutter-based calibration [43] is widely used since a blackbody is large, heavy, and expensive. Therefore, a method that is inspired by object-based calibration can be achieved by leveraging different parts of the same body with similar emissivity for calibration.

Several factors causing errors still exist on the thermal camera in a real environment. The emissivity of a target object is one of the factors, which is the efficiency of the surface in thermal energy emission. The emissivity is a specific parameter dependent on a component of a substance. Mitchell et al. [44] reported the difference in the emissivity of human skin is 0.95 – 0.99, caused by the difference in the blood flow and skin color. This 4% difference leads to $\pm 0.25^{\circ}$C error in the measurement of skin temperature. In order to determine the emissivity, it is necessary to measure the temperature of an object whose emissivity is known, such as black body tape [45]. This issue based on the idea that the reference points share almost the same emissivity with the target have to be considered to measure human body temperature. Therefore, the method which is not affected by skin color is required because the body parts of the same person as the reference and the target are used. Moreover, in the environment which is in daily life, where blood flow does not change greatly.

However, the correction of the other factors is still challenging. Electrical noise and noise due to fluctuation in the scaling of thermal energy are the remaining major factors. To mitigate these effects, we essentially need high-end thermal cameras if we do not rely on any additional devices.

2.2 Related Work on Thermal Sensation

2.2.1 Thermal Sensation and Physiological Information

Originally, Fanger proposed the predicted mean vote (PMV) based on the 7-point scale [14]. The PMV calculates the thermal sensation from environmental factors, such as temperature, humidity, wind speed, and heat radiation, and factors on the human body, such as metabolism and the amount of clothing. However, the PMV requires the measurement of input values around the target person, which is difficult in a real-world setting. In addition, the PMV assumes a static environment, where the thermal sensation is not dynamic. The actual thermal sensation labeled by the occupant is called TSV. It has been reported that the PMV does not always agree with the TSV [46], [15]. This is because the TSV reflects individual preferences, of which estimation is inherently difficult.

Many methods have been proposed to estimate thermal comfort. Brager et al. [47] proposed a method that switches the model depending on the season. This study indicates that it is difficult to estimate the TSV by simply sensing only environmental factors, such as indoor temperature. According to Fanger [14], thermal sensation changes with environmental factors, such as temperature and humidity, as well as human factors, such as metabolic rate and clothing. Takada et al. [48] proposed a model to estimate the TSV in a dynamic environment from the core temperature and average skin

temperature of each part of the human body obtained using the 7-point method [49]. These studies indicate that body temperature is an essential factor for determining human thermal sensation. There is also a method [50] that detects discomfort owing to the air-conditioned environment with high accuracy using only heart rate fluctuations. It demonstrates that the TSV can be estimated using detailed heart rate variability. Heart rate variability is one of the indexes of relaxation. It is calculated by frequency analysis of the heart rate signal. Electrodermal activity is an index of sweating [51]. There are many studies that use it as an index of the TSV [52], [53], [54]. Electrodermal activity is observed actively when the user is in hot environments.

2.2.2 Estimation of Thermal Sensation Using Wearable Sensors

Barrios et al. [16] estimated the TSV using a method based on machine learning from the heart rate data measured by a chest strap sensor. Although a wristband sensor, used as a heart rate sensor, is also considered in this study, the measurement error is larger than that of a chest strap owing to motion artifact. Because of this reason, they also implied that the heart rate measured by the wristband sensor could not be used as a feature value. Liu et al. [55] proposed the use of a wearable sensor on the ankle in addition to an arm-mounted sensor and the chest strap sensor. An estimator was constructed for individuals by collecting physiological information in their daily lives. While this method can estimate the TSV with consideration to the subject's difference, it requires many wearable sensors. In addition, it is difficult to introduce such sensors in a real-world setting, such as offices or schools, from the perspective of invasiveness.

Studies have also been conducted to estimate the PMV using wearable sensors. Hasan et al. [56] proposed a method to calculate the PMV by estimating a person's metabolic rate from their heart rate and acceleration measured by a wristband sensor. However, it is known that the PMV does not necessarily correspond to the user's thermal sensation in the real-world setting because it does not assume a dynamic environment [46], [15]. Thus, air conditioning control or the evaluation of the environment according to the PMV is not always appropriate. In addition, most methods use only statistical values, such as the average value of the measured physiological information for the past several minutes or hours. Moreover, they do not consider the fluctuations of physiological information. Deep learning-based methods, such as long short-term memory (LSTM), become key solutions to perform an estimation corresponding to a dynamic environment. This is because the deep learning-based methods can extract time-series fluctuation information from the time-series samples of each physiological information compared with other traditional approaches.

2.2.3 Estimation of Thermal Sensation Using Thermal Images

In recent studies, a thermal camera has been used to estimate the TSV because it can non-invasively measure skin temperature, such as the face temperature distribution. Li et al. [57] proposed a thermal camera network to assess thermal comfort in multi-occupancy spaces. Our method can be utilized in

such an environment with occupants wearing a wrist-worn device. Ranjan et al. [6] suggested that energy consumption in a building can be reduced by estimating the TSV using a thermal image. Furthermore, Aryal et al. [20] estimated the TSV by combining thermal images and other biosensors. Their method improved the estimation accuracy using features, including statistical values, e.g., average, maximum, and minimum, in a dynamic environment where the indoor temperature gradually changes. However, their work focused on comparing the pairs of sensors as feature values. Therefore, the evaluation was conducted using a continuous dataset from static subjects. In other words, they did not assume an actual environment, which includes the situation where a thermal image cannot be captured, as mentioned in Section 4.1. Additionally, this study did not evaluate how the previous measurement values affect the estimation of the TSV. The performance can be improved by focusing on how time-series data from only a wristband sensor improves the estimation accuracy in an actual environment.

Cosma et al. [58] extracted the temperature of multiple parts of the face from a thermal image. They collected learning data in a steady environment with an air conditioner (AC) at a fixed temperature setting and evaluated the estimation accuracy in a dynamic environment where the temperature setting is gradually changed. However, because they intentionally reproduced extreme environmental change, slight changes in room temperature in the actual environment are not evaluated. Specifically, in these studies, the maximum and minimum room temperatures were changed to approximately 30°C and 20°C, respectively. Moreover, they did not assume a dynamic environment in which the surrounding environment changes intermittently owing to a person's movement to another location and the gradual adaption of the human body to the environment. For such reason, thermal cameras can not deal with dynamic environments. To deal with the problem, time-series information from a wristband sensor helps to complement the lack of information only using thermal cameras.

2.2.4 Deep Learning-based Approaches

Various research has shown that deep learning-based techniques better estimate TSV than other traditional machine learning techniques. Hu et al. [59] showed their deep transfer learning-based approach was superior to other machine learning approaches such as Naive Bayes, Support Vector Machine, Decision Tree, Multi-layer perceptron, and k-Nearest Neighbors. However, this study transfers the knowledge trained by open datasets, which do not consider the dynamic situation. Morresi et al. [60] proposed TSV estimation models based on multiple machine learning algorithms, including Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). They showed these neural network-based approaches using time-series data effectively worked on estimating TSV in a dynamic environment. However, this study collected the dynamic data under the environment where the air temperature was drastically changed in the range wider than 10°C in the same trial. By contrast, our method focuses on more realistic environments where the air temperature is controlled normally.

Especially, LSTM, which is a special kind of RNN, is capable of learning long-term dependencies.

Basic RNN has a chain of repeating neural network modules with a single tanh layer. Instead of such a simple structure, LSTM is specialized to keep information longer than simple RNN as described in Ref. [61]. Many researchers found that LSTM works well on a large variety of applications, including TSV estimation.

The deep learning-based approach is suitable for a realistic environment because it can extract more information. Transfer learning is one of the approaches to combine multiple deep neural networks. Therefore, it can also be used to combine the opportunistic features and the time-series features. To extract time-series-based information and opportunistic information, the estimators are trained separately in the transfer learning framework. The information is concatenated after the extractions using those deep neural networks. The eventual estimator can be trained after the concatenation.

2.2.5 Applications Using Estimated TSV

The TSV is a useful metric for evaluating the energy efficiency of buildings. Cóstola et al. [62] proposed a novel performance indicator to describe trends in the energy performance of buildings based on TSV. Giamalaki et al. [63] conducted a questionnaire survey on the thermal experience of elderly people in their residences. The TSVs they collected indicate greater dissatisfaction with the thermal environment in heating than in the cooling season. Jain et al. [64] proposed an optimization framework for the comfort-energy trade-off. They realized the framework using an application that can record the change of the AC setting. TSV estimation can directly suggest comfortable AC settings without collecting such records. Enhancement of TSV estimation performance can make the suggestion more accurate.

Optimized spot-type personalized air-conditioners have been attracting more attention recently because they can conserve building energy. Zhu et al. [65] proposed the cooling efficiency index for personalized air-conditioners. Their system utilizes computational fluid dynamics and Fanger's thermoregulation model to simulate heat loss from the human body. Moreover, Qiao et al. [66] proposed a novel personal cooling vest incorporated with phase change materials and ventilation fans. Also, appropriate HVAC control enhances productivity [3] and learning performance [4], [67] in office buildings and schools. TSV estimation can also be used to optimize the trade-off between thermal sensation and energy consumption for such a spot-type air-conditioner.

2.3 Related Work on Data Balancing

2.3.1 SMOTE-based Data Balancing

In order to deal with the imbalanced datasets, researchers proposed data augmentation methods. The basic strategy is pre-processing the data to increase its size and diversity [68]. SMOTE is a predominant data augmentation for classification. Chawla et al. [17] showed the advantages of this approach compared to other alternative sampling techniques on several real-world problems using several clas-

sification algorithms. Because of the advantage, methods derived from the SMOTE are proposed [69], such as Adaptive Synthetic sampling approach (ADASYN) [70] and Borderline-SMOTE [71]. The SMOTE-based extensions replace the original interpolation procedure with other more complex ones, such as clustering and probabilistic functions.

Furthermore, filtering extensions after SMOTE are proposed, such as SMOTE + Tomek [72] and SMOTE + ENN [73]. To clarify the boundaries between classes, they remove unnecessary samples from the dataset after the data augmentation. They are kinds of undersampling techniques. In this desertion, to focus on evaluating the data augmentation of time-series data, the combination with undersampling is out of scope. However, those undersampling techniques can be applied after the data augmentation.

For regression problems, SMOTER [18] and SMOGN [74], which is an extension of SMOTER with Gaussian noise, are proposed. These methods extend SMOTE for regression problems by using the relevance function representing the density of the training data. In Chapter 5, we extend this algorithm for regression problems based on SMOTER, which is one of the SMOTE-based extensions for regression problems. It separates the minor cases from the distribution by a user-defined threshold. After the separation, it generates new cases based on weighted averages between pairs in minor cases.

2.3.2 Data Balancing for Time-series Classification

For time-series classification, some balancing methods were proposed by generating the synthetic sample. TS_SMOTE [22] is an extension of SMOTE designed for time-series data. It introduced DTW in the time-series merging algorithm in the augmentation. On the other hand, OHIT [75] was proposed as an oversampling method for the imbalanced time-series classification. OHIT is different from the state-of-the-art oversampling algorithms because it generates structure-preserving synthetic samples. These methods are based on generating samples using the feature of time-series categorized as the same class. In contrast, data balancing methods for time-series regression, which can not use classified time-series samples, are not fully investigated.

2.3.3 Learning-based Data Balancing

Another approach for data augmentation is known as generative models [76] which are based on deep learning, such as Generative Adversarial Networks (GANs) [77] and Variational Autoencoders (VAEs) [78]. These learning-based approaches can generate synthetic time-series data and augment training datasets effectively [79]. The algorithms model the real data distribution P_r by learning a distribution P_θ parameterized by θ. The data is generated by learning a function g_θ which transform a noise with gaussian distribution Z such that $P_\theta \approx g_\theta(Z)$. The approaches generate realistic values in several domains, such as computer vision and cybersecurity [80].

However, the generative models need training, which means the generative model may overfit common values when we use an imbalanced dataset. To deal with the problem, it needs a specialized loss

function, which makes the model more complicated, or data balancing before the training. Therefore, the data augmentation method based on the algorithmic approach, which does not need training, is required.

Chapter 3

Smartphone Thermal Camera Correction Using a Wristband Sensor

3.1 Introduction

Thermal cameras are widely used to monitor the temperature [81] because they can measure temperature distributions quickly without physical contact. For this reason, they are used in various fields such as animals [82, 83], agriculture [84], buildings [85] and humans. For example, thermal cameras are applied to humans for anomaly detection such as heatstroke [86] and infection [87]. Human skin temperature is also used for estimating various psychological states such as stress and thermal comfort [5, 6, 88]. Their applications are still limited due to the limitation of single-point measurement with some wearable devices with skin temperature sensors are available (e.g., E4 wristband [10] and Embrace 2 [11]),

Recently, thermal cameras have been widely used because low-cost and miniaturized models are available on the market. For example, FLIR ONE [12] is a low-cost thermal camera attached to a smartphone. Thanks to such smartphone thermal cameras, we can measure our skin temperature anytime, anywhere. However, the accuracy of the low-cost thermal camera is insufficient to monitor the skin temperature compared with the high-end models.

Thermal cameras are categorized into two types: those with cooled infrared detectors and those with uncooled infrared detectors [13]. The performance of the cooled detectors is much higher than the other, although they are bulky and expensive due to the cooling apparatus. Therefore, the smartphone thermal cameras are with uncooled infrared detectors. The uncooled infrared detector element converts its temperature rise to electric signals. The temperature of the object is principally calculated from the signals and their emissivity. The measured temperature greatly fluctuates in accordance with the parameters configured by the user, the camera's body heat, the efficiency of the element, and the

packaging method [13]. By these effects, the measurement error of the smartphone thermal camera is larger than the high-end one. The error range of FLIR ONE is ±3°C or ±5%, which is larger than the range of human skin temperature changes in daily life. This is not enough for various healthcare applications. For example, Abdelrahman [5] reported the nose temperature downs 1.49 °C when participants read a literary piece accompanied by increasing the cognitive load.

To overcome this problem, we propose a dynamic offset correction method for thermal images of faces captured by smartphone thermal cameras. We fully utilize the key feature of the thermal cameras: the measurement fluctuation is mainly caused by the offset, which is common in all the pixels in a single thermal image. In other words, we can measure the difference of temperature correctly between any pair of pixels in the same thermal image, even using the smartphone thermal camera. There is a method that uses the temperature reference fixed in the environment. It is actually used in airports etc. However, our method does not require any fixed devices and achieves accurate mobile thermal measurement. Our method combines thermal images with a reliable absolute temperature obtained by a wristband sensor based on the above feature. First, we obtain a thermal image including a reference point (a wrist or a palm) of which absolute temperature is measured by a wearable device (a wristband sensor). Second, we estimate the offset at the reference point by comparing the temperature measured by the thermal camera and the wristband sensor. Finally, our method corrects temperature in the thermal image by adding the offset to all the pixels.

Since we assume to measure the skin temperature, the reference point should be any point, which has an emissivity as large as the face. We use the wristband sensor as an accurate thermometer, which is widely used in daily life. Since the wristband sensor covers the measurement point, the thermal camera cannot directly capture the temperature of the same point. Therefore, we define the reference point as the point which has a high correlation with the point measured by the wristband sensor. In this chapter, we use a palm or a wrist as the reference point and compare the performance through the real experiment.

To evaluate our method, we conducted real experiments with eight subjects in an indoor air-conditioned environment. The mean absolute error and the standard deviation of the face temperature measurement error decrease by 21.0% and 68.3%, respectively, and Pearson's correlation coefficient increase by 138%, highlighting the effectiveness of our method. In addition, we show the combination of our method and the correction by continuous averaging observations improves the evaluation values. Finally, in order to evaluate the influence on our method due to environmental temperature change, we conducted an additional experiment in the cool environment and reveal the limitation with respect to the ambient temperature.

The main contributions of this paper are summarized as below:

- We propose an offset correction method for thermal images of faces measured by a smartphone thermal camera by using a reference point combined with the specific feature of thermal cameras without any fixed devices.

- We compare the two reference points and discuss their features through the real data with 876 samples collected from eight subjects.
- We investigate the limitation of our method on ambient temperature and show it can be used in the indoor air-conditioned environment.

3.2 Smartphone Thermal Camera Correction Method

3.2.1 Overview

Infrared detectors built in thermal cameras composed of a number of detector elements having different signal responsibilities. To measure quantitatively, the nonuniformity is corrected using a method called nonuniformity correction (NUC) [41]. Because of NUC, there is little difference in the accuracy between detector elements. Therefore, the error of a thermal image is a common offset over all pixels.

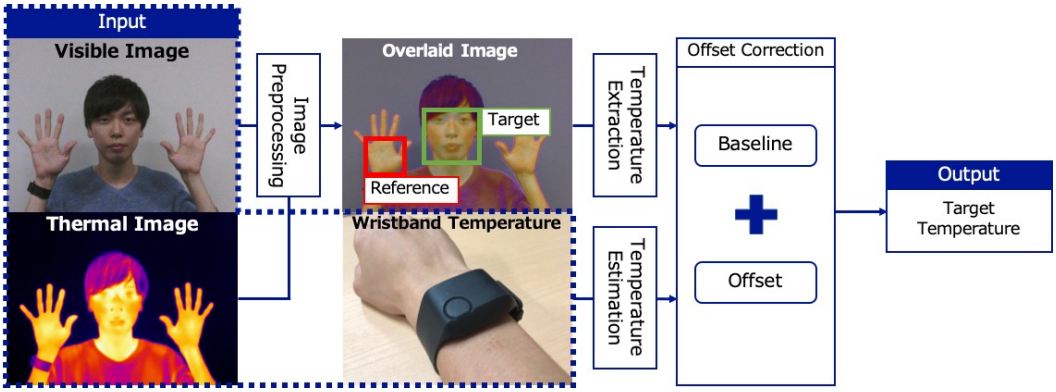


Figure 3.1: Overview of the offset correction

We correct the offset by setting a reference point in a thermal image. Figure 3.1 shows an overview of our method. We assume a user wears a wristband sensor such as an E4 sensor to measure her wrist skin temperature in real-time. The input is wrist temperature measured by the wristband, a visible image, and a thermal image. We extract the skin temperature of body parts from the thermal image by using image processing techniques. To do this, we apply pre-processing to match each pixel in the visible image with the temperature in the thermal image. Based on linear regression, the reference point temperature is estimated from the temperature of the wrist. By comparing the reference point temperature measured by the wristband sensor and the thermal image, we estimate the offset. Finally, we obtain the target point temperature by adding the offset to the thermal image.

We note that the limitation of wristband thermometers is low reliability when people sweat. However, our target applications include comfort level estimation for air conditioning and psychological stress estimation in daily life. We assume most of these applications are used in air-conditioned indoor

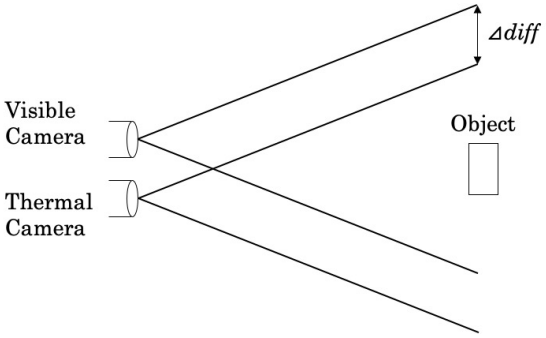


Figure 3.2: Position difference of visible and thermal cameras

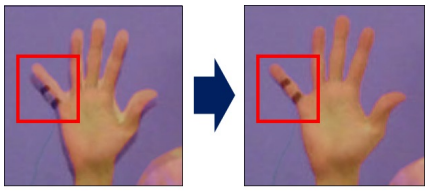


Figure 3.3: Result of overlapping

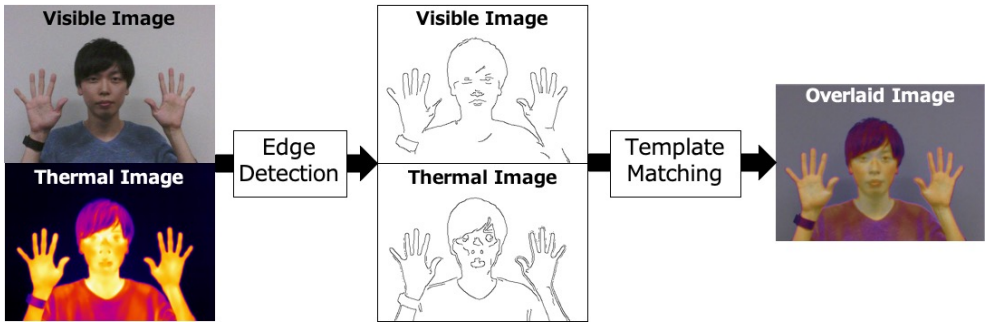


Figure 3.4: Image pre-processing steps

environments where people seldom sweat. Therefore, we are able to rely on a wristband thermometer as a reference.

3.2.2 Image Pre-processing

As shown in Figure 3.2, the visible image and the thermal image have slightly different views due to the difference between camera positions. This means, for example, the palm pixels in the visible image are not exactly equivalent to those in the thermal image. Therefore, if we recognize a target region (e.g., a palm) in the visible image and then obtain the temperature distribution of the area from the thermal

Figure 3.5: Palm region extraction

image, the distribution wrongly contains the temperature of different parts. Figure 3.3 shows that this problem greatly affects temperature extraction, especially for a small part such as a fingertip and a nose. Also, the position difference has a large impact on feature extraction from the distribution, such as the minimum value and the maximum value. Therefore, we apply image pre-processing to match each pixel in the visible image with its correct temperature in the thermal image.

In pre-processing, the key idea is that both images share almost the same edges. Figure 3.4 shows the steps in the pre-processing. First, edge detection using the Canny algorithm [89] is performed on both images. Smoothing based on a Gaussian filter is applied to both images after edge detection. Finally, Normalized Cross Correlation [90] is used for template matching.

3.2.3 Reference Point Temperature Extraction

The temperature of any pixel in the visible image can be extracted from the thermal image after the pre-processing. In this paper, we propose two reference points close to the wristband. One is a palm, and the other is a wrist around the wristband. In the following sections, we describe the temperature extraction for each reference point.

Palm Temperature Extraction

We extract a palm from a visible image based on skin color as shown in Figure 3.5 because it is difficult to detect hands from thermal images. This is because the skin is easily cooled by the atmosphere in the cold environment, and it is assimilated into the background in the thermal image.

We use an object detector using OpenCV based on Haar-like feature [91] to detect a palm from a visible image. The position of the palm can be obtained as a rectangle. Since we can obtain the area, including other parts such as a background and clothing, we need to further extract the palm region. For this purpose, we use an approach for human skin detection proposed by Tan et al. [92]. We use the green value of RGB color space and the saturation value of HSV color space [93] to obtain a squared histogram of color values in the face. We regard a pixel whose distance from the center of the smoothed histogram [94] is within three times of the standard deviation as a skin pixel. We extract the temperature of the skin pixel and calculate the average of the temperatures as a palm temperature

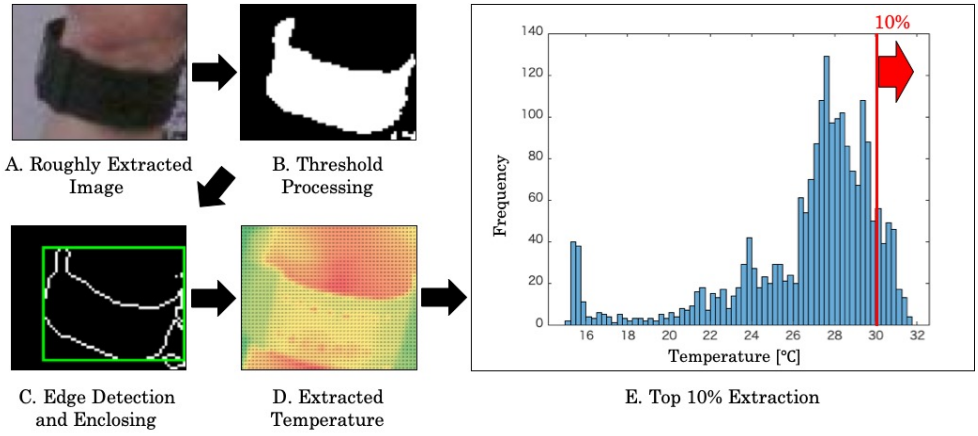


Figure 3.6: Flow of wrist temperature extraction

$T_{\text{palm\_meas}}$. The definition is given as

$$T_{\text{palm\_meas}} = \frac{\sum_{(x,y)\in P} T(x,y)}{|P|}, \tag{3.1}$$

where (x, y) is a coordinate in a visible image and P is a set of the extracted skin pixels. Also, $T(x, y)$ is the temperature of (x, y) and $|X|$ is the number of elements in the set X. P is defined as

$$P = \{(x,y)|(x,y) \in R_{\text{hand}} \cap C_{\text{hand}}\}, \tag{3.2}$$

where R_{hand} is a set of coordinates in the detected palm rectangle and C_{hand} is a set of coordinates whose colors are regarded as skin.

Wrist Temperature Extraction

We extract a wrist from a visible image based on the wristband color. Different from the palm detector implemented in OpenCV, we need to implement the wristband detector. In this paper, we manually extracted rough positions of the wristband. However, we note that this may be easily achieved by attaching a special marker or collecting training data. The flow of the wrist temperature extraction is shown in Figure 3.6. First, to highlight the wristband, we conduct threshold processing based on the hue value from HSV color space. Next, we conduct edge detection, enclose the detected edges in a rectangle, and extract temperatures in the enclosed box from corresponding pixels of thermal images. Finally, we extract wrist temperature $T_{\text{wrist\_meas}}$ by calculating the average of top 10% in the temperature distribution. $T_{\text{wrist\_meas}}$ is defined as

$$T_{\text{wrist\_meas}} = \frac{\sum_{(x,y)\in W} T(x,y)}{|W|}, \tag{3.3}$$

Figure 3.7: Face skin detection

where W is a set of the wrist coordinates which is defined as

$$W = \{(x, y)|(x, y) \in R_{\text{wrist}} \cap C_{\text{wrist}}\}. \tag{3.4}$$

In the above equation, R_{wrist} is a set of coordinates in the enclosed box of the wristband and C_{wrist} is a set of coordinates whose temperature is in the top 10%.

3.2.4 Reference Point Temperature Estimation

As we mentioned earlier, thermal cameras cannot capture the temperature of the point measured by the wristband sensor since it is covered by the wristband sensor itself. Therefore, we need to correlate the wristband temperature with the temperature of the reference points. For this purpose, we construct regression models given as

$$y = a + bx, \tag{3.5}$$

where x is an explanatory variable which is a temperature reported by the wristband sensor, y is a target variable as a temperature of the reference point (T_{est}) which is a palm ($T_{\text{palm\_est}}$) or a wrist ($T_{\text{wrist\_est}}$). a and b are parameters determined by training data.

3.2.5 Target Part Extraction

In this chapter, we select faces as the target part because the face temperature is used in many applications [6, 8, 5, 24]. We note that our method can also be applied to other target parts. We use the face detector of OpenCV based on Haar-like feature [91]. Since we can obtain a face position as a rectangle, including a background, we need to extract the temperatures of facial skins from the thermal image. For this purpose, we use an approach for human skin detection [92] to the visible image and extract from the corresponding point in the thermal image. Figure 3.7 shows an example of skin detection. We extract face temperature T_{face} by calculating the average of face pixels. The definition is given by

Table 3.1: Specification of FLIR T540 and FLIR ONE

	FLIR T540	FLIR ONE 2
IR Sensor Resolution	464×348	160×120
Accuracy	± 1°C or ± 1% (10–35°C)	± 3°C or ± 5% (0–35°C)
Thermal Sensitivity	0.04°C	0.15 °C

$$T_{\text{face}} = \frac{\sum_{(x,y)\in F} T(x,y)}{|F|}, \tag{3.6}$$

where F is a set of coordinates in the detected face rectangle.

3.2.6 Offset Correction

We obtain the corrected temperature T_{corr} of the target by adding the offset C to the temperature T_{target} of the target in the thermal image as below.

$$\begin{aligned} T_{\text{corr}} &= T_{\text{target}} + C \\ C &= T_{\text{est}} - T_{\text{meas}}. \end{aligned} \tag{3.7}$$

The offset C is calculated for the reference point (either a palm or a wrist) by subtracting the temperature T_{meas} measured by the thermal camera from the temperature T_{est} estimated by the regression. Namely, T_{est} is either $T_{\text{palm\_est}}$ or $T_{\text{wrist\_est}}$ and T_{meas} is either $T_{\text{palm\_meas}}$ or $T_{\text{wrist\_meas}}$.

3.3 Evaluation

3.3.1 Evaluation Settings

For evaluation, we used FLIR ONE 2 and an E4 wristband as a smartphone thermal camera and a wristband sensor, respectively. We also used a high-end thermal camera, FLIR T540, for the ground truth. The specifications of FLIR T540 and FLIR ONE 2 are shown in Table 3.1. The smartphone thermography has a lower resolution and accuracy than the high-end one, but both of them have small thermal sensitivity, and thermal differences can be accurately measured. E4 wristband measures the temperature at the sampling rate of 4 Hz. We use the mean temperature per minute for the evaluation. The accuracy of the temperature sensor is 0.2 °C within $36-39$ °C and its thermal sensitivity is 0.02 °C. It is accurate enough to calculate a reference point temperature.

To evaluate our method in detail, we conducted two experiments at different ambient temperatures. All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Graduate School of Information Science and Technology, Osaka University

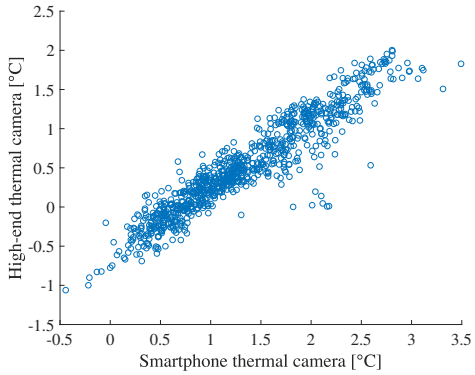


Figure 3.8: Relation between temperature difference (face – wrist) measured by high-end thermal camera and by smartphone thermal camera

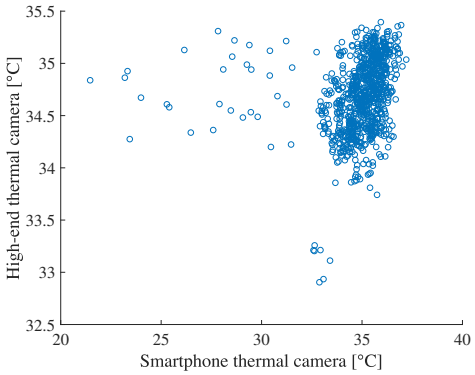


Figure 3.9: Relation between face temperature measured by high-end thermal camera and smartphone thermal camera

(201811). First, we collected the real data from eight male subjects aged the twenties for six hours. In the experiment, we captured both visible and thermal images seven times with a ten-second interval every 30 minutes. During the experiment, the subjects were asked to wear E4 wristbands and work as usual in the laboratory. They sat in front of the thermal cameras with the palms of both hands facing toward the cameras without any overlap. After removing some images which are incorrectly captured, 876 samples (i.e., pairs of the visible and thermal images) were collected in total. The maximum and the minimum numbers of samples per subject are 105 and 77, respectively. The average room temperature of the experiment was 26.18 °C and the standard deviation was 0.80 °C. The highest and the lowest air temperatures in the room were 30.6 °C and 21.9 °C, respectively.

In addition, we conducted an experiment in a roughly controlled cool environment to evaluate the influence on our method due to environmental temperature change. We used the same equipment as the previous experiment and collected data from two subjects in their early twenties for 14 hours for each subject over two days. Both of them participated in the previous experiment. We capture the images seven times with a ten-second interval every 30 minutes. In total, we collected 448 samples. The average temperature of the room was 17.16 °C and standard deviation was 1.14 °C.

Hereafter, we denote the former experiment as “lab” and the latter experiment as “cool.” We note that the room temperature of the lab experiment is clearly warmer than the cool experiment.

3.3.2 Confirmation of Thermal Camera’s Feature

We found that thermal cameras are suitable for measuring the temperature difference in the same image, even if a low-cost smartphone thermal camera. To confirm it, we compare the temperature difference between the face and around wristband measured by the high-end thermal camera with the

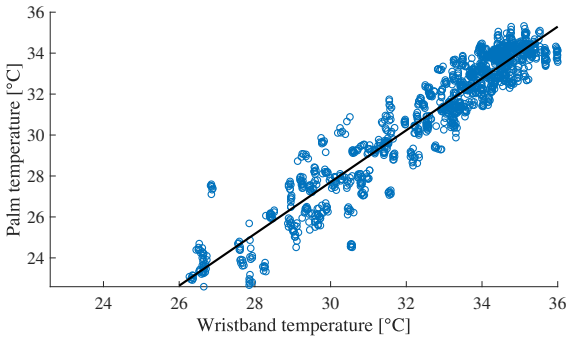


Figure 3.10: Relation between temperature measured by wristband and palm temperature

same measurement by the smartphone thermal camera. Figure 3.8 illustrates the relation between them in the lab experiment showing a high correlation coefficient was 0.94. However, Figure 3.9 shows a low correlation between the measurement by the high-end thermal camera and the smartphone thermal camera. The correlation coefficient was 0.21, which is totally different from the temperature difference. From the results, we see smartphone thermal cameras can still capture temperature differences, which is comparable with the high-end ones. We also note that the temperature difference measured by the smartphone thermal camera tends to be slightly higher than the high-end one. We describe it in section 3.3.4.

3.3.3 Reference Point Temperature Estimation

We evaluated the performance of the temperature estimation of the reference points by linear regression. The purpose of the experiments is to build models to estimate the reference point temperature from the wristband temperature. Therefore, we used the thermal images captured by the high-end thermal camera to collect the training data.

The regression functions are sufficient to understand the difference temperature between palm/wrist and wristband as shown in figure 3.10 and figure 3.11. Figure 3.10 and figure 3.11 show the relation between the reference points' temperature measured by the high-end thermal camera and the temperature measured by the wristband in both experiments described in section 3.3.1. The palm is used as a reference point in figure 3.10, and the wrist is used as a reference point in figure 3.11. The line on the figures is the regression lines, and they also show the estimation formulas. The regression function for estimating the palm temperature is given as

$$T_{\text{palm\_est}} = -10.2505 + 1.2651\ T_{\text{wristband}}. \tag{3.8}$$

The mean absolute error was 0.7598 °C.

On the other hand, the regression function for estimating the wrist temperature is given as

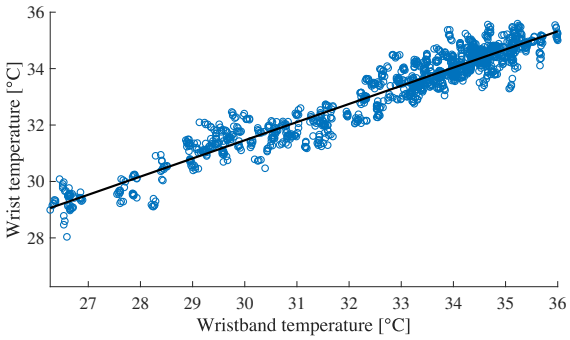


Figure 3.11: Relation between temperature measured by wristband and temperature around wristband

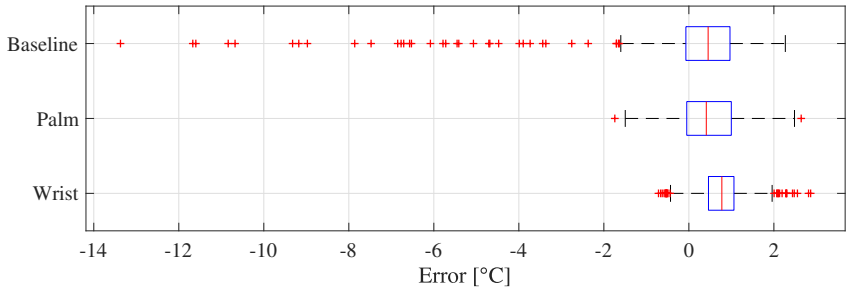


Figure 3.12: Error distributions of face temperature

$$T_{\text{wrist\_est}} = 12.1492 + 0.6437\ T_{\text{wristband}}. \tag{3.9}$$

The mean absolute error was 0.3460 °C.

From the above results, we see that the wrist temperature can be estimated more accurately than the palm. This is natural because of the closeness to the point measured by the wristband sensor. However, the errors of the two methods are much smaller than the error of the smartphone thermal camera. It shows both are valid as reference points. We further investigate the performance of our method based on the above regression functions in the next section.

3.3.4 Dynamic Offset Correction

Figure 3.12 shows the error distributions of the baseline (without correction), our palm referenced method, and our wrist referenced method. We see that the distribution approaches 0 in the palm referenced method while the error dispersion becomes smaller in the wrist referenced method. Our methods remove the large error in the baseline completely. The result is important to capture changes and abnormalities of body temperature. It is because that applications need to be sensitive to capture

Table 3.2: Result of dynamic offset correction

Method	MAE	SD	CORR
Baseline	0.93	1.59	0.21
Palm referenced	**0.73**	0.72	0.44
Wrist referenced	0.79	**0.50**	**0.49**

the change of the body temperature between a person at different times. Also, when we capture a single person in a thermal image for abnormality detection, such as fever screening, we have to capture the body temperature without large error in an one-shot thermal image.

Also, Table 3.2 shows the mean absolute error (MAE), the standard deviation of the error (SD), and the Pearson's linear correlation coefficient (CORR) of the baseline and our methods (including the baseline). It is obvious that both of our methods reduce MAE and SD compared to the baseline. The palm referenced method achieves 0.20 °C smaller MAE than the baseline, which is also 0.06 °C smaller than the wrist referenced method. This is because the resolution of the smartphone thermal camera is much smaller than the high-end one. We found that the average temperature in a small region (e.g., a wrist) tends to become higher when the resolution is high. The low resolution leads to difficulty in capturing the correct temperature distribution in a small region. This is described in section 3.3.5 for more detail. On the other hand, the wrist referenced method achieves a smaller SD than the other. This is because the accuracy of the reference point temperature estimation for the wrist is higher than the palm. Finally, the CORR of our methods is remarkably higher than the baseline. It means our methods achieve higher linearity, which indicates the accuracy of sensor correction.

From the above results, the palm referenced method is suitable for accuracy, while the wrist referenced method is suitable for high precision. In practice, accuracy may be more important for many applications. For example, the applications for cognitive load estimation [5] and thermal comfort estimation [6] need to capture temporal changes and differences among different persons. However, if we can mitigate the low-resolution effect of the smartphone thermal camera, the wrist referenced method may achieve better in terms of both accuracy and precision.

3.3.5 Effect of Ambient Air Temperature

The above experiments have shown the effect of the reference point temperature difference on the performance. Table 3.3 shows the SD of the correction error in each environment. The SD is 0.22 °C larger than the lab environment.

Figure 3.13(a) – Figure 3.13(d) show the distributions of the top 10% extraction of the wrist temperature. Because the distribution of a smartphone thermal camera includes an offset error, the minimum values of the distributions are aligned. Figure 3.13(a) shows the distribution captured by the high-end thermal camera in the lab environment, and the line is the average (33.68 °C). Figure 3.13(b) shows the distribution captured by the smartphone thermal camera in the lab environment, and the line is

Table 3.3: Difference in standard deviation due to environmental change

Method	Lab	Cool
Baseline	1.63	1.76
our method	0.50	0.88

the average (33.67 °C). Figure 3.13(c) shows the distribution captured by the high-end thermal camera in the cool environment, and the line is the average (30.64 °C). Figure 3.13(d) shows the distribution captured by the smartphone thermal camera in the cool environment, and the line is the average (30.24 °C). In the warmer temperature environment, the two distributions show the same trends, and there is the least difference between the averages. However, in the cooler temperature environment, the two distributions show different trends, and the average of the high-end one is higher than the smartphone one. The reason is the difference in resolution. With the decrease of the temperature, the distribution range becomes larger. Because the low-resolution thermal cameras cannot capture the detailed distribution, the temperature of the narrow region measured by the smartphone thermal camera tends to be low.

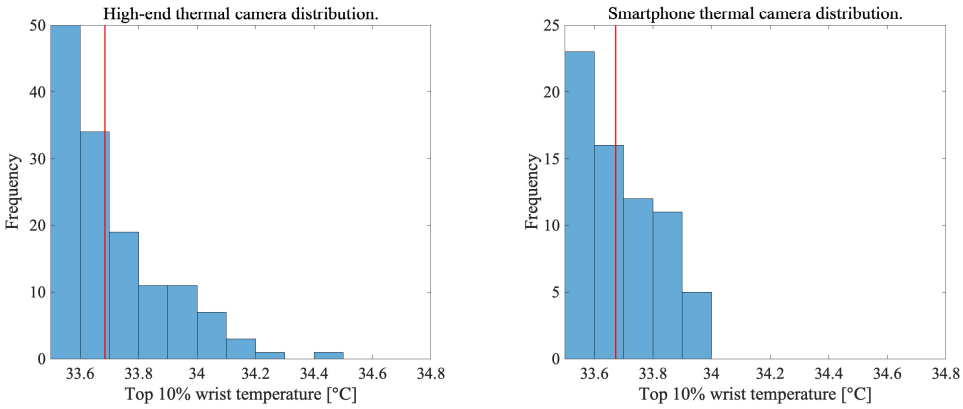
Figure 3.14 shows the difference between the temperature around the wristband and the face. The vertical axis is measured by the smartphone one, and the horizontal axis is measured by the high-end one. The difference in the lab experiment tends to be larger than the difference in the cool experiment. We assume that we can capture a similar distribution of the face temperature distribution compared with the wrist because the region of the face is much larger than the region around the wrist. Under this assumption, the difference in Figure 3.14 means the wrist temperature measured by a smartphone camera is lower than the ground truth. Therefore, there is a limitation of the temperature variation of the reference point in our method. The wristband temperature is the criterion of the environmental temperature.

Figure 3.15 shows the relation between the number of data and the error of our method. The data in the dataset used for calculating the SD are extracted by setting the lower limit of the temperature measured by the wristband. When the cool data is mixed at 32.60 °C, the SD suddenly increases. The wrist temperature border is shown in Figure 3.15 as the blue line. The mean room temperature when the bordering temperature was recorded was 20.87 °C. For this reason, we consider that the wrist temperature needs to be 32.60 °C or higher in order to utilize our method with a small error.

3.3.6 Performance in Continuous Measurement

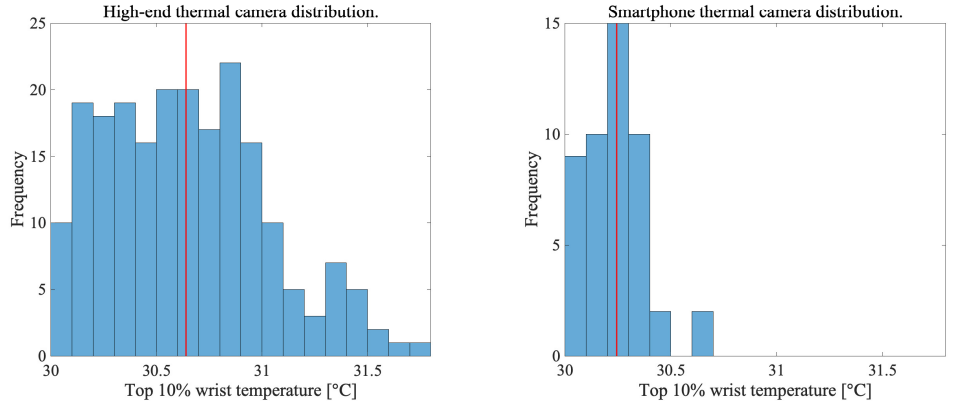
To see the performance of our method in continuous measurement, we also conducted an experiment in the laboratory environment. The subject is a male aged twenties. We continuously recorded the images of a subject for 10 minutes.

Figure 3.16 shows the change of the temperature of the baseline, the ground truth, and our method



(a) High-end, Lab

(b) Smartphone, Lab



(c) High-end, Cool

(d) Smartphone, Cool

Figure 3.13: High-end and smartphone thermal camera distribution in lab and cool environment.

using a palm temperature over time. Even if we remove some large fluctuations of the baseline as outliers, there still remains a larger error than our method. This is because the error characteristic of smartphone thermal cameras is unpredictable and sometimes biased due to environmental factors such as the camera body temperature itself. Meanwhile, our method largely improves the measurement results following the ground truth.

3.3.7 Comparison with Continuous Measurement Correction

The accuracy of smartphone thermal cameras may be mitigated by conducting multiple measurements to remove outliers. To see the effect of such simple outlier detection, we captured both visible and thermal images seven times with a ten-second interval every 30 minutes. Then, we remove outliers from continuous measurement by defining a suitable threshold. Figure 3.17 shows an example of the

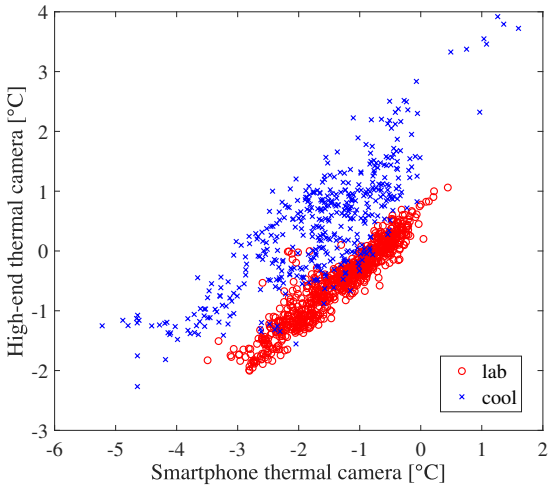


Figure 3.14: Difference between temperature around wristband and face

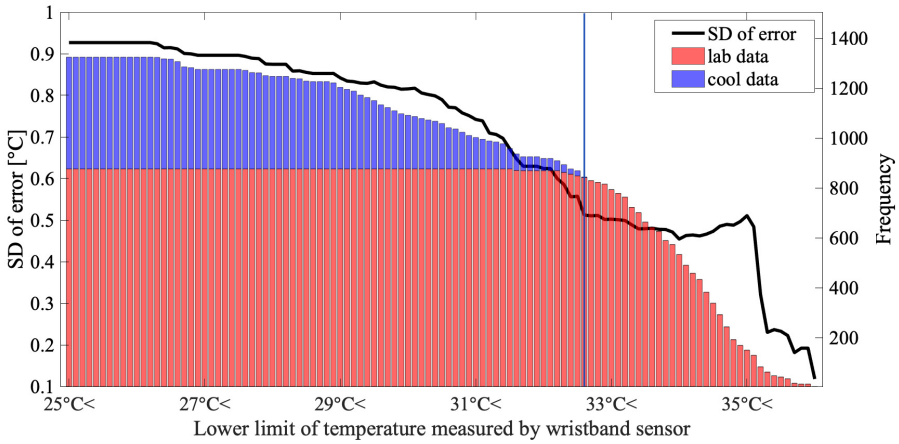


Figure 3.15: Relation between wristband measured temperature and SD of correction error

outlier removal process. The black line is the mean of the seven measurements, and the red lines are the thresholds. The thresholds are defined as the temperatures which are higher or lower than the SD of the seven continuous measurements.

Table 3.4 shows MAE, SD, and CORR for the dataset where the above outlier removal was applied. We see that all the results are improved by the outlier removal. The MAE of the baseline is especially improved by the outlier removal, which is slightly better than our method. However, the SD and CORR of the wrist referenced method are much better than the baseline even with the outlier removal. This result indicates that our method greatly improves the performance of smartphone thermal cameras

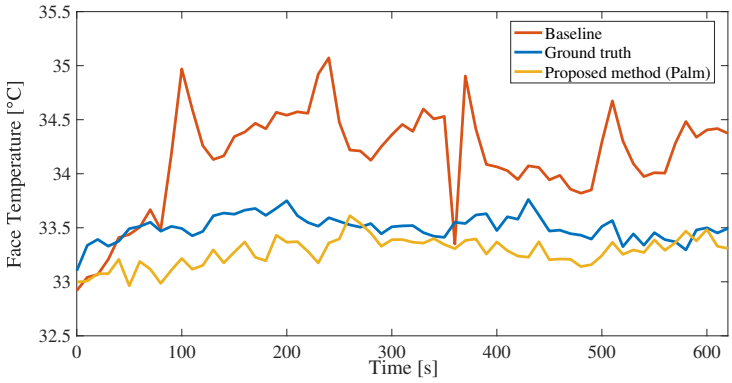


Figure 3.16: Temperature change over time

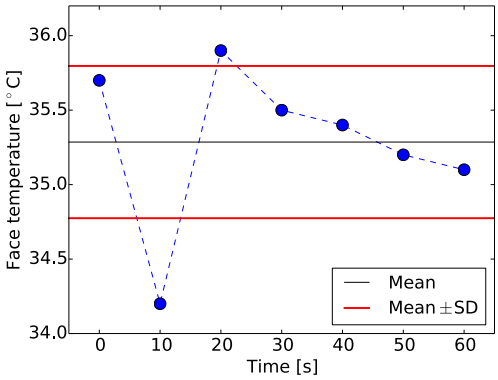
Table 3.4: Effect of outlier removal

Method	MAE	SD	CORR
Baseline	0.69	0.79	0.37
Palm referenced	0.71	0.79	0.46
Wrist referenced	0.76	0.47	0.56

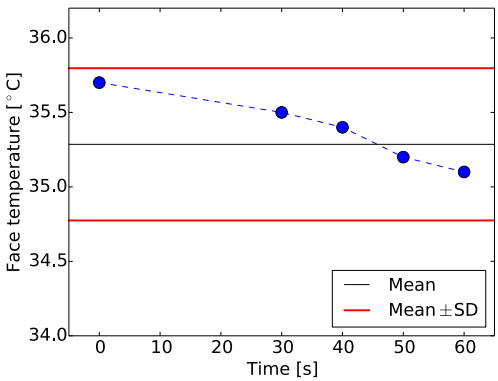
especially in terms of the ability to follow the changes of the ground truth. Therefore, we have confirmed that our method is also useful for continuous measurement.

3.4 Conclusion

In this chapter, we presented dynamic offset correction for a smartphone thermal camera using a wristband sensor for low cost and accurate temperature monitoring. The design of our method is based on the key feature that the measurement fluctuation of the thermal cameras is due to the offset, which is common in all the pixels in a single thermal image. Our method estimates the temperature of the reference point by regression from the wristband temperature measurement. We selected a palm and a wrist for the reference points for comparison. Through the real experiment with 876 samples from eight subjects, we confirmed that our method remarkably improved the accuracy on three evaluation metrics: the mean absolute error, the standard deviation of the error, and Pearson's linear correlation coefficient. The limitation of our method depends on the variance of reference point temperature distribution. The experiment in a cool environment shows the variance is larger when the temperature measured by a wristband sensor is lower. The wrist temperature needs to be 32.60 °C or higher in order to fully utilize our method. This result indicates our method is effective in normal indoor ambient temperature.



(a) Raw continuous measurement



(b) After outlier removal

Figure 3.17: Example of removal defining thresholds

Chapter 4

Combining Time-Series and Opportunistic Sensing for Dynamic Thermal Sensation Estimation

4.1 Introduction

The estimation of a person's thermal sensation enables the realization of an appropriate thermal environment. However, it is challenging to always estimate their thermal sensation correctly in indoor environment. This is because a person's thermal preference dynamically changes, owing to the spatial non-uniformity of the environment or a change in the weather [95], [96]. They cause a fluctuation of a person's ideal air environment when a person moves to different spaces, such as different rooms in the same building or different seats in the same room [95]. Therefore, the same air quality control does not always provide the same environmental comfort. For these reasons, it is necessary to obtain an individual's dynamic thermal sensation to monitor the air-conditioned environment.

Many studies have been conducted to estimate thermal sensation. The American Society of Heating, Refrigerating, and Air-Conditioning Engineers' 7-point thermal scale [97] has been widely used as the metrics of human thermal sensation; the seven levels range from -3 to +3 (Cold, Cool, Slightly cool, Neutral, Slightly warm, Warm, Hot). The actual thermal sensation labeled by the occupant is called the thermal sensation vote (TSV) using this metric. To estimate the TSV, many studies have been conducted recently based on machine learning techniques using physiological data measured by wearable sensors [16], [55]. In these studies, a chest strap sensor was used as the heart rate sensor instead of a wristband sensor, owing to the limitation of measurement accuracy of the wristband sensor. However, it is difficult to introduce the chest strap sensor in a real-world setting, owing to the user's burden. Although it can continuously monitor users' physiological data, a wristband sensor is more preferable in daily use. However, heart rates measured by wristband sensors are less accurate than chest heart rate monitors, which means estimation of thermal sensation by a wristband sensor

is difficult as reported in Ref.[16]. To solve the above problems, we combine a thermal camera with a wristband sensor for thermal comfort estimation. Our method estimates user's TSV based on his physiological data and face temperature distribution by machine learning. Even if user's thermal image cannot be captured, estimation is performed using the wristband sensor. To enhance the accuracy, we leverage the recent estimation results obtained by the combination of the thermal camera and the wristband sensor. As a preliminary study, we collected data of wristband sensors and thermal cameras from 15 male subjects for 128 days in different seasons in the laboratory environment, capturing 1,476 thermal images in total. We evaluated the effectiveness of our thermal comfort estimation to reduce energy consumption of ACs. From the result, we confirmed that our method can estimate thermal comfort with F-measure of 0.85, allowing shifts to the neighboring classes.

As the latter part of this chapter, we focus on the TSV estimation in more realistic situation. A study have been conducted to estimate the TSV by measuring the body surface temperature using a thermal camera as a non-invasive sensor to obtain more detailed information [6]. In this study, the body surface temperature around the face is measured by a thermal camera and added to the feature set. Consequently, accurate estimation for a static environment has been achieved. However, this study did not evaluate how people's thermal sensation change in the transient environment. Furthermore, in the studies [16], [20], [59], [58], [98] that address a dynamic environment, the dataset used for evaluation has a temperature range as wide as $8 - 10^{\circ}$C. In real air-conditioned environments, such temperature change does not happen. Therefore, a method is required for TSV estimation in a dynamic environment, which is more realistic.

However, because the change in physiological information in an air-conditioned environment is small, the estimation is challenging. For this problem, we propose a TSV estimation method suitable for a realistic environment using deep learning. The key idea is to combine time-series information with opportunistic thermal images obtained by thermal cameras to capture a slight change of physiological information. In the real-world setting, users may move around to different spaces. Under such a situation, it is difficult to continuously measure the environmental data, such as air temperature and relative humidity, around the user. Therefore, our method uses only opportunistic environmental data around the user measured by the environmental sensor attached with a thermal camera or a spot-type HVAC system. Deep learning-based methods are known for the capability to extract information, which is difficult to design for humans. In our method, physiological time-series information measured by a wristband sensor is combined with the temperature distribution of the face, measured by a thermal camera, to reflect a slight physiological response to the current environment for the estimation of the TSV. The method is based on the idea that the thermal sensation significantly correlates with heat-loss from the skin surface in the dynamic environment [99]. To reflect the heat-loss in TSV estimation, we input the time-series information from the wristband sensor. In this study, we assume that a wristband sensor periodically measures physiological information, and the face temperature is measured opportunistically by a thermal camera installed in the environment, such as on the door

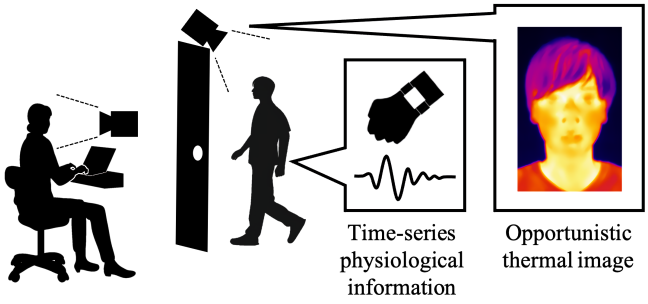


Figure 4.1: Environmental setting

or desk, as shown in Figure 4.1. When the thermal camera measures the temperature, our method combines it with the physiological information obtained by the wristband sensor to estimate the TSV based on machine learning. To combine the opportunistic and time-series features, we build our method based on a deep transfer learning. The basic idea is as follows. First, we train the partial estimator using time-series data only. Second, to combine the knowledge extracted from the time-series data with the opportunistic data, our method transfers the partial network of the estimator for the time-series data to the input of the latter combined TSV estimator. Finally, the transferred network is combined with the opportunistic data as the input to the latter combined TSV estimator. The detail of the model is described in Section 4.2.

For evaluation, we collected data from 21 subjects for spring, summer, autumn, and winter. In total, we collected 1686 TSV inputs. We note that our study is conducted in a challenging environment because most of the data are collected in an air-conditioned environment to investigate the potential of improving human comfort in the real-world setting. Therefore, we exclude the data labeled as discomfort from the collected data for the evaluation, which accounts for 0.8% of total reports of TSV. As a result, most of the TSVs labeled by participants are 0 (Neutral). This makes the estimation more challenging compared with the existing state-of-the-art studies, which divide the occupant's thermal state into comfort and discomfort [19], [20], [21]. Because such an imbalanced dataset is used, we also propose a data balancing method for the time-series data in the regression problem to prevent biased learning. As a result, the macro-averaged F1-score is improved by 5.8% compared with the baseline estimation without considering the time-series information. Furthermore, we evaluate the length of time required for the consideration of past physiological information and show that consideration of the previous 10 minutes is effective in TSV estimation. The main contributions of this study are summarized below.

- We separately construct the estimator that uses the physiological information acquired as a time-series and the estimator that combines the time-series information and opportunistically measured information. The evaluation result shows that our method is more effective than a

Table 4.1: Features of a thermal camera used in machine learning

#	Feature	Description
f_1	T_{face}	Whole face temperature
f_2	T_{forehead}	Forehead temperature
f_3	$T_{\text{cheek\_r}}$	Right cheek temperature
f_4	$T_{\text{cheek\_l}}$	Left cheek temperature
f_5	T_{nose}	Nose tip temperature
f_6	T_{mouth}	Lips temperature
f_7	T_{max}	Max value in facial pixel
f_8	T_{min}	Min value in facial pixel
f_9	T_{var}	Variance of facial pixel

baseline method.

- In a regression problem, we design a balancing method for an imbalanced dataset, including time-series data. We demonstrate the effectiveness of the method by evaluating its effect on TSV estimation performance.
- By changing the duration of the past physiological information used in our method, we reveal the duration of the past time that should be considered in TSV estimation.

4.2 TSV Estimation Method

4.2.1 Reliability-based Combination of Thermal Camera and Wristband Sensor

Our method estimates thermal comfort based on features measured by a thermal camera and a wristband sensor by machine learning. We assume large rooms such as offices and schools and low-cost thermal cameras are distributed in the environment. As low-cost thermal cameras, we assume thermal cameras such as FLIR ONE [12] that can be attached to smartphones. Our method switches between a combined model and a wrist-based model according to the availability of the thermal images as illustrated in Figure 4.1. The features extracted from the thermal images are shown in Table 4.1. Some of them are used in Ref. [6] while the others are selected based on our analysis using our dataset. From a thermal image, we extract skin temperatures of different facial parts by applying image processing techniques. For the face detection, we use OpenCV face detector based on haar-like feature [91]. For the detected bounding box of the face, we define the relative positions of each part as shown in Table 4.2. Then, to precisely extract skin pixels in the thermal image, we overlay the visible image and the thermal image as in our previous work [100] and apply a skin detection method [92] to the overlaid image. Finally, we extract the features from the skin pixels.

The features obtained by a wristband sensor is shown in Table 4.3. We measure skin temperature (WT), heart rate (HR), and electrodermal activity (EDA). To consider changes over time, we calculate

Table 4.2: Rectangular calculation formula showing the position of each part of the face

Part	x,	y	Width,	Height
Forehead	$X + 7W/16$,	$Y + H/4$	$W/8$,	$H/8$
Right cheek	$X + W/5$,	$Y + 9H/16$	$W/6$,	$H/4$
Left cheek	$X + 19W/30$,	$Y + 9H/16$	$W/6$,	$H/4$
Nose	$X + 7W/16$,	$Y + H/2$	$W/8$,	$H/8$
Mouth	$X + 3W/8$,	$Y + 3H/4$	$W/4$,	$H/8$

Table 4.3: Features of a wristband sensor used in machine learning

#	Feature	Description
f_{10}	WT	Wrist temperature
$f_{11} - f_{13}$	$WT_{\text{mean\_}n\text{min}}$	WT mean between {1, 5, 10} minute
$f_{14} - f_{16}$	$WT_{\text{diff\_}n\text{min}}$	WT difference between WT and {1, 5, 10} minute before
f_{17}	HR	Heart rate
$f_{18} - f_{20}$	$HR_{\text{mean\_}n\text{min}}$	HR mean between {1, 5, 10} minute
$f_{21} - f_{23}$	$HR_{\text{diff\_}n\text{min}}$	HR difference between HR and {1, 5, 10} minute before
f_{24}	EDA	Electrodermal activity
$f_{25} - f_{27}$	$EDA_{\text{mean\_}n\text{min}}$	EDA mean between {1, 5, 10} minute
$f_{28} - f_{30}$	$EDA_{\text{diff\_}n\text{min}}$	EDA difference between EDA and {1, 5, 10} minute before

the means in 1, 5, and 10 minutes and the difference from 1, 5, and 10 minutes ago. The AC mode is also used as a feature f_{31} representing heating (+1) or cooling (-1). When the thermal camera cannot capture user's face, we use a wrist-based model by using available features. To increase the accuracy, our method also uses the latest estimation result with the thermal images considering the elapsed time as a weight representing its reliability.

$$C_{est}(t) = a^{t-t_{prev}} C_{prev} + \left(1 - a^{t-t_{prev}}\right) C_{wrist}(t), \tag{4.1}$$

where C_{prev} is the latest estimated TSV using all features at time t_{prev}. $C_{wrist}(t)$ is an estimated TSV using the features without the thermal image at time t. In this paper, a is empirically set to 0.9.

4.2.2 Overview of TSVNet

We propose TSVNet that combines a wristband sensor and a thermal camera by deep learning-based approach. An overview of TSVNet is shown in Figure 4.2. The values measured by the wristband sensor

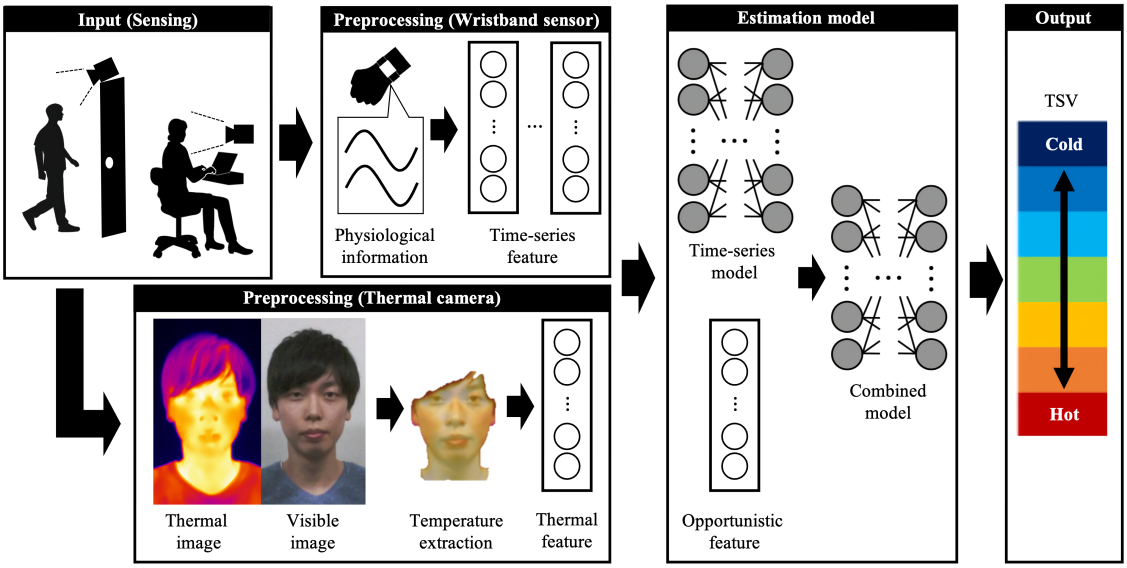


Figure 4.2: Overview of TSVNet

are used as time-series information on the physiological response to slight changes in the air-conditioned environment. Furthermore, TSVNet occasionally captures the face temperature distribution using a thermal camera. The face temperature distribution reflects the physiological response to the current air-conditioned environment [6]. To combine the time-series information and current information, we build multiple estimators based on deep learning and combine them to construct a combined estimator. In the following, we construct an estimator based on time-series physiological information measured by a wristband sensor and an estimator based on occasionally-measured physiological information and environmental information. Finally, we describe a method to construct the combined estimator.

4.2.3 Preprocessing

Measurement with Wristband Sensor

We extract three types of time-series data from the wristband sensor. First, the heart rate W_{hr} is extracted as an index of discomfort. It is known that the heart rate increases in uncomfortable environments, such as hot or cold temperatures [16]. Second, we use the wrist temperature W_{temp} because local thermal sensation affects the whole-body sensation, as mentioned in existing studies [101], [102], [103]. Third, we use electrodermal activity W_{eda}, which is known as an index of sweating [104]. For estimation

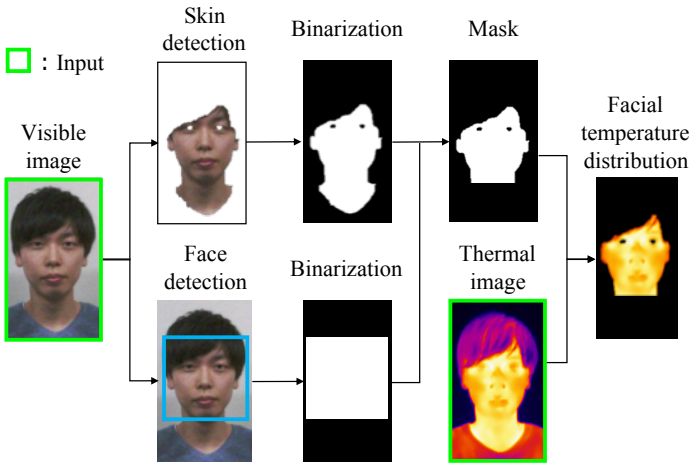


Figure 4.3: Flow of facial temperature distribution extraction

at time t, we use continuously measured physiological data from time $t-N$ to t as follows:

$$\boldsymbol{W}_{hr}^{t} = \{W_{hr}^{t}, ..., W_{hr}^{t-N}\}, \tag{4.2}$$

$$\boldsymbol{W}_{temp}^{t} = \{W_{temp}^{t}, ..., W_{temp}^{t-N}\}, \tag{4.3}$$

$$\boldsymbol{W}_{eda}^{t} = \{W_{eda}^{t}, ..., W_{eda}^{t-N}\}, \tag{4.4}$$

where W_X^t is the measured value of the physiological data X obtained by the wristband sensor at time t. This information is the input to the estimator at time t. We use an average value for 1 minute as the measured value at time t.

Measurement with Thermal Camera

We assume that a thermal camera captures the temperature distribution of the user's face when the user is in front of it. We note that most thermal cameras capture visible images and thermal images. Therefore, we also assume that the user in the thermal image is identified by state-of-the-art identification methods using visible images [105]. For the extraction of only the facial temperature from the temperature distribution around the face, preprocessing is performed, as shown in Figure 4.3. First, a mask image is generated from the visible image, extracting the facial skin surface. The mask image is generated using a pixel-wise logical AND operation between the skin detection [92] and the face detection using the haar-like feature [90]. The skin detection is based on eye detection, which determines a pixel as the subject's skin based on the color of pixels around the eyes. Therefore, as mentioned in the literature [92], the method is robust to variations in illumination, background image, camera characteristics, and ethnicity. Face detection is based on the haar-like feature implemented in OpenCV [90]. Finally, only the facial temperature distribution is extracted from the thermal image by

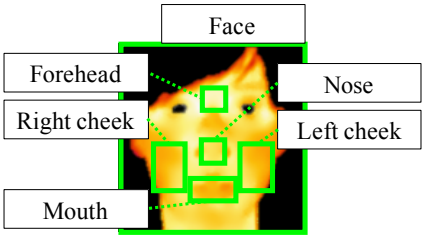


Figure 4.4: Position of each part of the face

the generated mask image. Because different cameras capture the visible and thermal images, there is a small shift between the positions of the subject in each image. For masking without the shift, it is necessary to map each pixel in the visible image to one of the pixels in the thermal image. Therefore, we apply an overlay method [106] based on the edges that typically appear in both images. Using the above processes, we accurately extract the skin temperature distribution around the face.

As an input to the estimator, features are extracted from the masked temperature distribution. As shown in Figure 4.4, we extract the temperature around the forehead, right cheek, left cheek, nasal tip, and mouth. We also use the maximum, minimum, average, and variance of the facial temperature distribution as statistical information. The respective values at time t are expressed as T^t_{fh}, $T^t_{r\_chk}$, $T^t_{l\_chk}$, T^t_{ns}, T^t_{mth}, T^t_{max}, T^t_{min}, T^t_{ave}, and T^t_{var}, hereafter. Each temperature is calculated as the average of the pixel values in each part defined by rectangles. We define the positions of the rectangles as listed in Table 4.2. In the table, X,Y,W, and H are the face's horizontal coordinate, vertical coordinate, horizontal width, and vertical width, respectively, and each coordinate is the upper left vertex coordinate of the rectangle.

4.2.4 Building Combined Estimator

To combine the time-series and opportunistic information for estimation, we construct a combined estimator based on transfer learning. The configuration of the combined estimator is shown in Figure 4.5. We apply standardization to all the features extracted by the sensors before they are input to the estimator.

The structure of the time-series-based estimator TS_{part} is shown in the upper part of Figure 4.5. The input data, which is acquired by the wristband sensor as time-series data, is defined as below:

$$\boldsymbol{W}^t = \{\boldsymbol{W}^t_{hr}, \boldsymbol{W}^t_{temp}, \boldsymbol{W}^t_{eda}\}. \tag{4.5}$$

The estimator is configured using one LSTM layer and two fully connected (FC) layers. The number of nodes is decided based on the number of features input to the model. Because the features input to the LSTM layer is three time-series, an extremely large number of nodes may cause overfitting. Our preliminary experiment found that the performance decreased when the number of nodes was

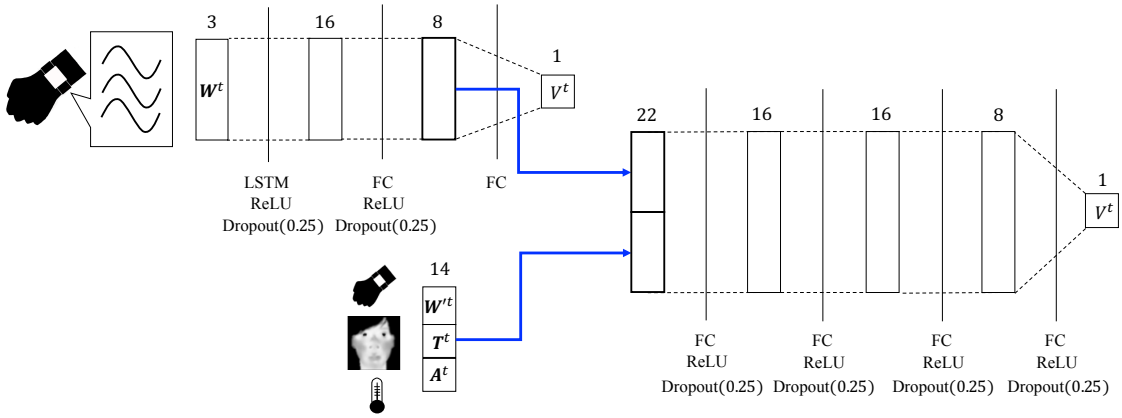


Figure 4.5: Estimator configuration

more than 16 because of the overfitting. In other words, it was found that fewer nodes than 16 are not powerful enough for the estimation. This means the network has not learned the relevant features in the training data. In this work, the number of nodes is decreased as few as possible for the generalization while maintaining the performance. Therefore, the number of LSTM output nodes is set to 16. For the generalization of the model, the number of nodes is reduced as the layer becomes closer to the output. For this reason, we set the number of the following nodes to eight. For consideration of the nonlinearity, the ReLU function is used as the activation function for the LSTM layer's output with the dropout rate of 0.25. Dropout is one of the most effective and most commonly used regularization techniques for complex neural networks by randomly dropping out the nodes. Considering the trade-off between the generalization and fitting performance of the model, we determine the dropout rate is 0.25 for the whole model. The ReLU function is also used for the output of the fully connected layer. The dropout rate to prevent overfitting is also 0.25. We note that our estimator outputs TSV V_t as a numerical value by learning it as a regression problem. This is because training for classification does not consider continuity of relations between features and target values.

The opportunistic features are directly input to the combined estimator. The inputs are acquired by a wristband sensor, a thermal camera, and an environmental sensor. The data measured by the wristband sensor is defined as

$$\boldsymbol{W'^t} = \{W^t_{hr}, W^t_{temp}, W^t_{eda}\}. \tag{4.6}$$

Moreover, the feature based on the facial temperature measured by the thermal camera is defined as

$$\boldsymbol{T}^t = \{T^t_{fh}, T^t_{r\_chk}, T^t_{l\_chk}, T^t_{ns}, T^t_{mth}, T^t_{max}, T^t_{min}, T^t_{ave}, T^t_{var}\}. \tag{4.7}$$

In addition to the above physiological features, the room temperature A^t_{temp} and relative humidity A^t_{humid} at time t are defined as an environmental feature:

$$\boldsymbol{A^t} = \{A^t_{temp}, A^t_{humid}\}. \tag{4.8}$$

A hidden layer of 22 nodes is configured by connecting the first hidden layers of TS_{part} and the opportunistic features. The combined estimator is retrained as a regression problem which consists of four FC layers. The numbers of the nodes are also determined by the aforementioned idea and the result of a preliminary experiment. As a result, the numbers of the first, second, and third layers' output nodes are set to 16, 16, and 8, respectively. The ReLU function is used as the activation function for the FC layer output, and the dropout rate is 0.25.

4.3 Oversampling Imbalanced Data for Regression

The dataset collected in this study is imbalanced, owing to many responses of Neutral, as described in Section 4.4.4. If we train our estimator on such a dataset, it may be biased to output Neutral. In this section, we propose a data balancing method for a regression problem, including time-series data.

In this study, we train each estimator as regression problems. Therefore, we use SMOGN [74], which is a data balancing method for regression problems for oversampling opportunistic features. SMOGN combines SMOTER [18] and data generation using Gaussian noise. SMOTER is a regression extension of SMOTE [17], which is an oversampling method for classification. However, we cannot apply SMOGN to our regression problem with the time-series input because SMOGN does not support the generation of time-series data. Therefore, we conduct data balancing by combining SMOTER with TS_SMOTE [22], which is a time series data generation method for two-class classification.

The pseudocode of the main algorithm for data balancing is shown in Algorithm 1. First, based on the SMOTER algorithm, the threshold t_E of $\phi(y)$ is determined to separate the data into major and minor classes. $\phi(y)$ is a function that shows the uniqueness of the target value y. A larger value indicates a smaller number of samples. The data generation for minor classes is performed by dividing the median of y into data that is smaller and larger than $\tilde{y}$. The method for generating time-series data is shown in Algorithm 2. The k-nearest neighbors of the minor class sample *case* are extracted based on the distance calculated by dynamic time warping [107]. A new time-series dataset *new* is generated by TS_SMOTE for the neighbor n, randomly selected from the k-nearest neighbors. After the target number of samples ng is generated by repeating the above steps, the new generated dataset $newCases$ is returned.

4.4 Evaluation

4.4.1 Evaluation Setting for Reliability-based TSV Estimation

We collected data using a wristband sensor E4 [10] and a thermal camera FLIR T540 [108] from 15 male subjects for 128 days through one year. The sensors embedded in the E4 wristband are summarized in Table 4.4. We measure wristband's features using the sensors at the frequency listed in the table. We captured seven thermal images with a ten-second interval every 30 minutes and used the mean of the seven images as a thermal image at that time. The subjects also recorded their TSV

Algorithm 1 Main SMOTER algorithm

Input: $\mathcal{D}$ - A data set
t_E - Threshold
$\%o, \%u$ - Percentages of over- and under-sampling
k - Number of neighbors used in case generation
Output: $\mathcal{D}_{new}$ - A generated data set
1: $rareL \Leftarrow \{\langle \boldsymbol{x}, y\rangle \in \mathcal{D} : \phi(y) > t_E \wedge y < \tilde{y}\}$
2: $newCasesL \Leftarrow$ genSynthCases$(rareL, \%o, k)$
3: $rareH \Leftarrow \{\langle \boldsymbol{x}, y\rangle \in \mathcal{D} : \phi(y) > t_E \wedge y > \tilde{y}\}$
4: $newCasesH \Leftarrow$ genSynthCases$(rareH, \%o, k)$
5: $newCases \Leftarrow newCasesL \cup newCasesH$
6: $nrNorm \Leftarrow \%u$ of $|newCases|$
7: $normCases \Leftarrow$ sample of $nrNorm$ case $\in \mathcal{D} \setminus \{rareL \cup rareH\}$
8: **return** $newCases \cup normCases$

Table 4.4: Sensors on E4 wristband

Sensor	Frequency
Photoplethysmogram [bpm]	64Hz
Electrodermal activity [µS]	4Hz
Skin temperature [°C]	4Hz

Table 4.5: The number of reported values for each class

Class	Number of reports
-3 (cold)	11
-2 (cool)	43
-1 (slightly cool)	185
0 (neutral)	919
1 (slightly warm)	257
2 (warm)	55
3 (hot)	6

and comfort (Yes/No) every 30 minutes. After removing some samples with missing data, we collected 1,476 samples of thermal images with TSV labels in total. Table 4.5 shows the number of reported values for each class. The uncomfortable label is only reported in the extreme conditions (i.e. +3 and -3). To avoid over-fitting due to unbalanced samples, we apply SMOTE algorithm after removing classes +3 and -3 because the numbers of samples of these two classes were extremely small compared with the other classes.

We selected Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Random Forest (RF) for machine learning algorithms. The evaluation is performed by leave-one-person-out cross validation which does not use the data of the target subject for the training. We use precision and recall for evaluation metrics in the following experiments.

Algorithm 2 Generating synthetic cases

Input: $\mathcal{D}$ - A dataset
$\%o$ - Percentages of oversampling
k - Number of neighbors used in case generation
Output: $\mathcal{D}_{gen}$ - Generated new cases

$newCases \Leftarrow \{\}$
$ng \Leftarrow \%o/100$
for all $case \in \mathcal{D}$ **do**
  $nns \Leftarrow \text{KNN}(k, case, \mathcal{D} \setminus \{case\})$
  for $i \Leftarrow 1$ **to** ng **do**
    $n \Leftarrow$ randomly choose one of the nns
    $\alpha \Leftarrow$ randomly choose in $[0, 1]$
    $new[y] \Leftarrow min(case[y], n[y]) + \alpha|case[y], n[y]|$
    for all $f \in$ features **do**
      $pairs \Leftarrow \text{DTW}(case[f], n[f])$
      for $t \Leftarrow 1$ **to** $|pairs|$ **do**
        $(t_{new}, v_{new}) \Leftarrow \text{timePoint}(pairs[t]_{case}, pairs[t]_n, case[f], n[f], \alpha)$
        $new[f, t_{new}] \Leftarrow v_{new}$
      end for
    end for
    $newCases \Leftarrow newCases \cup \{new\}$
  end for
end for
return $newCases$
function timePoint$(t_a, t_b, ts_a, ts_b, \alpha)$
  $t_{new} \Leftarrow min(t_a, t_b) + (t_a + t_b)/2$
  $v_{new} \Leftarrow min(ts_a[t_a], ts_b[t_b]) + \alpha|ts_a[t_a] - ts_b[t_b]|$
  return t_{new}, v_{new}
end function

4.4.2 Feature Comparison

First, we evaluated the effect of different features. Table 4.6, Table 4.7, and Table 4.8 show the comparison between different feature sets and different machine learning algorithms. We compare three feature sets: wristband-based features, thermal camera-based features, and all features. Overall, RF performs better than the others in all methods although the precision and the recall of each method are less than 0.5. This result implies the estimation of TSV is challenging due to its subjective aspect. However, it is still acceptable even if the classification results are shifted to the neighboring classes in terms of the detection of overuse of electricity. For example, even if a *warm* state is classified as *slightly warm*, it is still an underestimate when the AC is heating. This means both cases are the states of energy overuse where the AC setting can be controlled to reduce energy consumption in winter. Therefore, such shifts of classification results are not critical.

From this point of view, we compared the three feature sets by allowing classification shifts to the neighboring classes. From the results shown in Tables 4.9,4.10, 4.11, we see RF performs well without

Table 4.6: Estimation result using features from wristband sensor

Method	Precision	Recall
LR	0.22	0.30
SVM	0.24	0.30
KNN	0.21	0.23
RF	0.25	0.32

Table 4.7: Estimation result using features from thermal camera

Method	Precision	Recall
LR	0.29	0.42
SVM	0.32	0.41
KNN	0.31	0.37
RF	0.32	0.41

Table 4.8: Estimation result using all features

Method	Precision	Recall
LR	0.30	0.43
SVM	0.32	0.41
KNN	0.31	0.38
RF	0.33	0.46

Table 4.9: Estimation result using features from wristband sensor allowing shifts to the neighboring classes

Method	Precision	Recall
LR	0.68	0.75
SVM	0.67	0.71
KNN	0.65	0.66
RF	0.73	0.73

Table 4.10: Estimation result using features from thermal camera allowing shifts to the neighboring classes

Method	Precision	Recall
LR	0.75	0.79
SVM	0.75	0.78
KNN	0.76	0.81
RF	0.78	0.82

Table 4.11: Estimation result using all features allowing shifts to the neighboring classes

Method	Precision	Recall
LR	0.75	0.78
SVM	0.76	0.84
KNN	0.76	0.82
RF	0.80	0.86

Table 4.12: Precision and recall for detection of energy overuse

Method	Precision	Recall
Our method	0.79	0.92
Wristband only	0.44	0.88

Table 4.13: Confusion matrix using combined model

		Predicted				
		+2	+1	0	-1	-2
Actual	+2	4	5	3	0	1
	+1	9	21	15	6	0
	0	12	49	93	14	1
	-1	3	3	20	1	0
	-2	0	0	1	1	0

Table 4.14: Confusion matrix using wristband-based model

		Predicted				
		+2	+1	0	-1	-2
Actual	+2	2	6	5	0	0
	+1	22	7	19	3	0
	0	33	26	106	4	0
	-1	3	2	20	2	0
	-2	0	0	2	0	0

thermal images in terms of F-measure (LR: 0.71, RF: 0.73). Also, RF performs well when we use a thermal camera only. In addition, the precision and the recall become better with the fusion of thermal images. Therefore, we have confirmed the effectiveness of combining the wristband sensor and the thermal camera.

4.4.3 Effect of Past Estimation Results

To see the effect of our method considering past estimation results, we compared our method with the wristband-based method which uses wristband features only. We collected test data on a day in winter from eight males. We collected a thermal image every 30 minutes for each subject and TSV labels every 10 minutes.

Figure 4.13 and Figure 4.14 show the confusion matrix of the estimation result using combined medel and wristband-based model, respectively. Also, Table 4.12 shows the average precision and recall. From the result, we see that our method achieves precision of 0.79 and recall of 0.92 by combining the wristband sensor and the thermal camera. From the above result, we confirmed the potential of our method to detect the overuse of electricity.

4.4.4 Evaluation Settings for TSVNet

We conducted experiments in two scenarios (scenario A and scenario B) to collect the TSVs in a laboratory environment. The subject can report a TSV within the range of [-3.5, 3.5] by moving the seek bar up and down on our smartphone application. In the following evaluation, we define the input TSV in seven levels: [-3.5, -2.5] as Cold, (-2.5, -1.5] as Cool, (-1.5, -0.5] as Slightly cool, (-0.5, 0.5) as Neutral, [0.5, 1.5) as Slightly warm, [1.5, 2.5) as Warm, and [2.5, 3.5] as Hot. In both scenarios, an E4 wristband [10] was used as the wristband sensor, and a FLIR T540 [108] was used as the thermal camera. For preprocessing, we compute the averages of these values every one minute to input the

Table 4.15: Number of TSV reports in each scenario

TSV	Scenario A	Scenario B
-3 (Cold)	11	0
-2 (Cool)	42	24
-1 (Slightly cool)	193	88
0 (Neutral)	766	249
1 (Slightly warm)	196	71
2 (Warm)	37	6
3 (Hot)	3	0

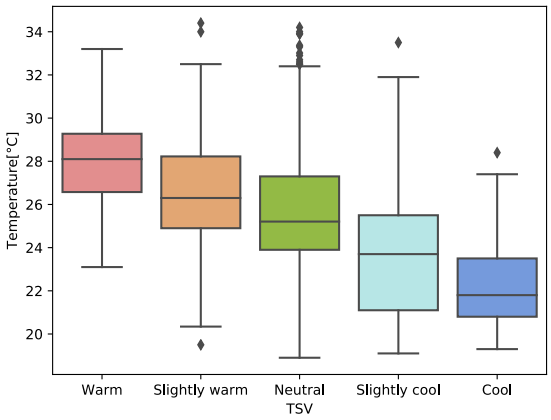


Figure 4.6: Distribution of air temperature in each class

estimator. In scenarios A and B, we did not specify the clothing of the subjects. In scenario A, the experiment was conducted in a static environment We collected TSV reports, thermal images, and visible images every 30 minutes in an air-conditioned room. During the experiment, the wristband sensor measured the heart rate, wrist temperature, and electrodermal activity. In this scenario, we collected data over 98 days from 15 males over all four seasons and obtained 1248 TSVs. The number of reported TSVs in each class is shown in Table 4.15. Neutral was the most frequently reported response; 3 (Hot) and -3 (Cold) were hardly reported because we controlled the room environment, as in daily life. In scenario B, the experiment was conducted in a dynamic environment, where the subject moved between two rooms every 30 minutes–1 hour. There is a temperature difference between 1°C and 5°C between them. In this scenario, we collected TSV reports, thermal images, and visible images every 10 minutes. For this experiment, data were collected over 25 days, including 438 TSV reports from ten males and two females. The number of TSVs in each class is also shown in Table 4.15. In this experiment, we assumed the actual environment with a slight temperature difference; therefore, 3 (Hot) and -3 (Cold) were not reported.

In total, we collected data for 123 days from 21 subjects. Out of 21 subjects, 6 participated in both scenarios. Each subject participated in data collection for 2–12 days. Figure 4.6 shows the relationship between the TSV and room temperature collected during the experiment. Because the numbers of Hot and Cold reports were small, we merged Hot (Cold) into the reports of Warm (Cool). As shown in Figure 4.6, the higher the room temperature, the more the TSV tends to approach Warm. However, the same room temperature may cause different reports, which makes TSV estimation difficult.

To confirm that our method estimates TSV without a bias to Neutral, which is most frequently reported, we evaluate our estimation as a classification problem by separating numerical TSVs into seven classes. However, because our dataset was collected in a realistic environment, the four extreme sensations, i.e., Cold, Cool, Warm, and Hot, are hardly reported. Therefore, we treated grouped classes into three categories, i.e., Cold, Cool, and Slightly Cool as Cool; Hot, Warm, and Slightly Warm as Warm; and Neutral. The evaluation is performed using leave one group out cross-validation to assess our method under the situation in which the target person's data is not included in the training data. We randomly divide the subjects into four groups. Then, three groups are used as the training data, and the remaining group is used as the test data. The estimation performance was compared based on the macro-averaged F1-score F_{ave} of each of three classes of Warm, Neutral, and Cool. It was calculated by the following formula based on the precision $P(c)$ and recall $R(c)$ of each class c.

$$\begin{aligned} F_{ave} &= \left(\sum_{c \in classes} \frac{2P(c)R(c)}{P(c)+R(c)} \right) / |classes|, \\ classes &= \{\mathrm{Warm}, \mathrm{Neutral}, \mathrm{Cool}\}. \end{aligned} \tag{4.9}$$

In addition, we calculate mean absolute error (MAE) as an evaluation metric for numerical estimation. For the following evaluation, we train the estimator ten times with a random selection of mini-batch and calculate an average of them. The evaluation metrics are the average of them.

4.4.5 Feature Importance

First, to confirm the same trend for features collected in our air-conditioned environment to those reported in existing literature [6, 16, 20], feature importance is calculated by inputting $\{\boldsymbol{W'^t}, \boldsymbol{T^t}, \boldsymbol{A^t}\}$, which are the features defined in Section 4.2.4, for estimation in a random forest. Figure 4.7 shows the result of calculating the importance of each feature. This value is calculated relative to the impurity that can be reduced by using a specific feature value when classifying the data. According to the result, the most useful feature in TSV estimation is the air temperature A_{temp}. Additionally, the cheek temperatures T_{r_chk} and T_{l_chk}, and the calculated average temperature T_{ave} of the face are also important. The wrist temperature W_{temp} is the most important among the values measured by the wristband sensor; however, the heart rate W_{hr} is less critical. We consider that this is influenced by the heart rate measurement accuracy of the wristband sensor, as reported in existing literature [16].

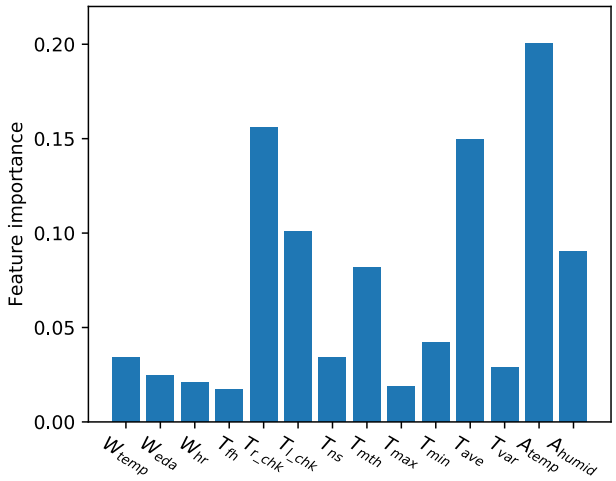


Figure 4.7: Importance of each feature for TSV estimation

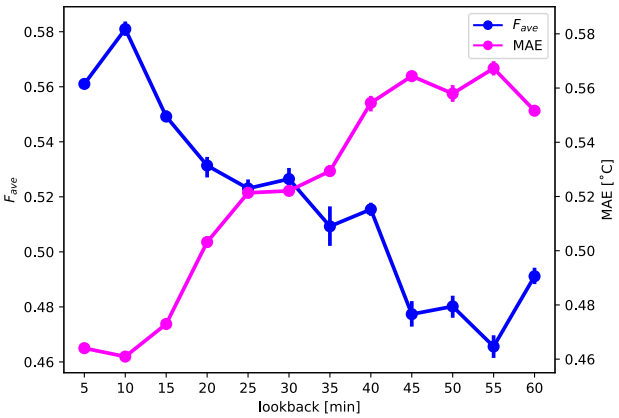


Figure 4.8: Relationship between lookback time and estimation performance

4.4.6 Effective Lookback Time for TSV Estimation

The physiological information measured by a wristband sensor for several minutes is used as a feature value. We evaluate the length N of the time (lookback time) used as the input to the time-series estimator TS_{part}. The estimation performance is shown in Figure 4.8, where the lookback time is changed between 5–60 minutes. The result highlights the most effective lookback time because the performance increases if sufficient features are included to explain the current TSV for the overall dataset.

The estimation performance improves as the lookback time increases to 10 minutes, and it is at its

Table 4.16: F_{ave} with or without data balancing

	TS_{part}	Our method
With balancing	**0.383**	**0.581**
Without balancing	0.300	0.462
%Δ	**27.7**	**25.8**

highest ($F_{ave} = 0.581$, $MAE = 0.461$) when the lookback time is 10 minutes for the overall dataset. The result shows that approximately 10 minutes of lookback on physiological information effectively enhance the estimation performance by reflecting the temporal transition of thermal sensation. We note that the result is represented by the dataset, including the data collected in a static environment. Therefore, positive impacts on F_{ave} and MAE possibly become more significant when a user frequently moves among rooms. On the other hand, we observed a drop in performance when the lookback time was longer than 10 minutes. If we make the estimator network more complex (e.g., increasing the number of nodes and adding layers), the performance can be increased. However, it exponentially increases the computational cost, such as time and memory for both learning and estimation. To mitigate an HVAC system's burden, we should reduce such costs. In the following, the lookback time is set to 10 minutes for evaluation.

4.4.7 Effect of Data Balancing

The time-series input dataset to TS_{part} is balanced by the method combining SMOTER and TS_SMOTE, as described in Section 4.3. In addition, SMOGN is applied to opportunistic features. In this section, we evaluate how the data balancing affects the estimation performance.

Table 4.16 shows a comparison of F_{ave} with and without data balancing. %Δ indicates the rate of F_{ave} improvement, owing to data balancing. The estimation performance of TS_{part} is the result that uses the partial feature extractor as an estimator. The result shows that the balancing method remarkably increase F_{ave} which are 27.7% and 25.8% for TS_{part} and our method, respectively. As a result, it is observed that the balancing method works effectively for the estimation performance because F_{ave} is improved in both cases.

4.4.8 Effect of Feature Combination

To evaluate how each feature works for the TSV estimation, we compare the estimation performance of our method with two estimators TS_{part} and OP_{part} using different features. The performance of TS_{part} indicates how only time-series features work for the estimation. In addition, to evaluate how only opportunistic features work, we use the combined estimator without the output of TS_{part}. OP_{part} uses opportunistic features only.

Table 4.17 and Table 4.18 show comparison of F_{ave} and MAE, respectively. As shown in the tables, our method shows the highest performance about F_{ave} and MAE for static, dynamic, and

Table 4.17: Comparison of TSV estimation F_{ave} with feature selection

Method	Static	Dynamic	All
Our method (Combined)	**0.568**	**0.585**	**0.580**
OP_{part} (Opportunistic)	0.539	0.540	0.552
TS_{part} (Time-series)	0.385	0.351	0.382

Table 4.18: Comparison of TSV estimation MAE with feature selection

Method	Static	Dynamic	All
Our method (Combined)	**0.472**	**0.431**	**0.461**
OP_{part} (Opportunistic)	0.479	0.451	0.472
TS_{part} (Time-series)	0.547	0.599	0.561

overall dataset. The result also shows that only time-series features do not work well compared with the opportunistic features. However, the combination of them increases the performance because our method shows the highest performance for each metric, i.e., F_{ave} and MAE. Therefore, the combination of the time-series and opportunistic features works effectively for TSV estimation.

4.4.9 Comparison with Existing Work

We implement the random forest-based method proposed in Ref. [20] for comparison with one of the existing state-of-the-art method in dynamic environment as a baseline. As described in the literature [20], [59], random forest exhibits excellent performance as a machine learning method used for thermal sensation estimation, particularly for small datasets. In this work, we assume the environment where we can measure time-series features by a wristband only. Therefore, the features input to the baseline method are recent measurement from all sensors and time-series based features proposed in Ref. [20] from wristband sensor. The time-series-based features are categorized into four different classes. The first one is raw values of measurements, while the second one is derivatives of the measurements. The third one is shapes of data streams, and the fourth is recent measurements. The features related to raw values of the measurements contain the minimum, maximum, average, standard deviation and median of the measurements in the 5-min window. The features related to the derivative include the minimum, maximum, average, standard deviation and median of the first derivative of the data stream. The features to capture the shapes of the data streams include the coefficients obtained by fitting a first degree (2 coefficients) and second degree (3 coefficients) polynomials to the measurements in the 5-min window. The features to capture recent measurements include the most recent measurement value, average value of last 10s, and average of the first derivative for the last 10s. This feature design is the same as Ref. [20].

Table 4.19 and Table 4.20 show the comparison of F_{ave} and MAE, respectively. $\%\Delta$ indicates the percentage of improvement achieved by our method compared with the baseline method labeled as

Table 4.19: Macro-averaged F1-score of TSV estimation

Method	Static	Dynamic	All
Our method	**0.568**	**0.585**	**0.580**
RF	0.528	0.550	0.548
%Δ	**7.0**	**6.4**	**5.8**

Table 4.20: Mean absolute error of TSV estimation

Method	Static	Dynamic	All
Our method	**0.472**	**0.431**	**0.461**
RF	0.508	0.471	0.498
%Δ	**-7.1**	**-8.5**	**-7.4**

Table 4.21: F1-score of each class for overall dataset

Method	Warm	Neutral	Cool
Our method	**0.344**	**0.775**	**0.623**
RF	0.344	0.707	0.593
%Δ	**0.0**	**9.6**	**5.1**

RF. As a result of the evaluation, for RF, F_{ave} for the static, dynamic, and overall dataset were 0.528, 0.550, and 0.548, respectively, and MAE for the static, dynamic, and overall dataset were 0.508, 0.471, and 0.498, respectively. For our method, F_{ave} for the static, dynamic, and overall dataset were 0.568, 0.585, and 0.580, respectively, and MAE for the static, dynamic, and overall dataset were 0.472, 0.431, and 0.461, respectively. Given that our method outperformed RF, the deep-learning-based approach to extract the time-series information improved the TSV estimation performance for all cases.

However, Table 4.21, which shows the F1-score of our method and RF for the estimation of each class, indicates that the classification of Warm and Neutral is more ambiguous than Cool and Neutral, even when using our method. One of the reasons is that the electrodermal activity used as an index of sweating did not work for the classification. This is because the sweating does not typically occur much during the actual daily Warm state, as shown in Figure 4.9. Therefore, it is suggested that the detection of sweating by the electrodermal activity is not sufficient to estimate the Warm state in daily life. Moreover, as observed from the distribution of the temperature on the right cheek in each class (see Figure 4.10), the body surface temperature drops significantly in the Cool state. However, in the Warm state, the body surface temperature did not rise considerably from Neutral. Therefore, the accurate classification between Neutral and Warm is challenging when only sweating and body surface temperature are considered. To address this limitation, we will collect more data to construct a personal estimation model that reflects the preference of each subject, such as clothes and temperature.

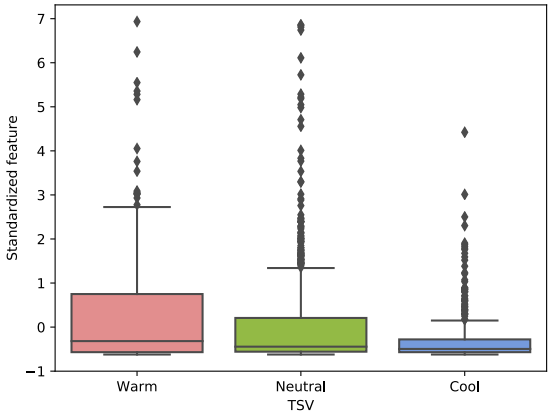


Figure 4.9: Distribution of electrodermal activity in each class

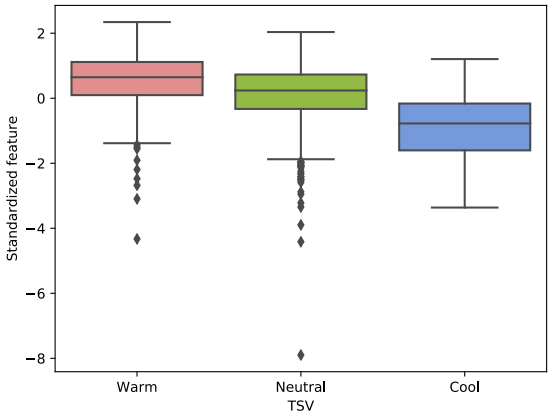


Figure 4.10: Distribution of right cheek's temperature in each class

4.5 Discussion

The estimation performance mentioned in Section 4.4.9 is relatively low compared with existing studies that address TSV classification [19], [20], [21], [58], [59]. However, we attempt TSV estimation under the challenging environment in which extreme temperature is not included. Owing to the above reason, the absolute error of our estimation result is high. However, one of our results exhibits the possibility of improving estimation performance by our method in both states, i.e., static and dynamic.

As an alternative to mitigate the dataset imbalance, we may collect data from "controlled" realistic environments that could occur in daily life. However, conducting experiments to collect physiological data comprehensively in a controlled environment often arises ethical and legal problems. In addition, it is difficult to eliminate the risk of causing problems on health completely. In the case of collecting

large amounts of training data, these constraints become a big issue to build an estimation model. In this study, we tackle the issue by using data balancing for physiological data collected in a wild environment.

Machine learning approaches are used to estimate a response variable, which can be estimated using explanatory variables from the measurable data. The estimation may not work well due to personal differences and differences that cannot be captured by the data used as explanatory variables. However, the estimator performance can be improved by adding explanatory variables and fitting the estimator to personal data. Although Ref. [109] and Ref. [110] show that the clothing level is an important factor for person's thermal sensation, we do not use the clothing level as a feature value. This is because we focus on applicability of thermal sensation estimation to real-world setting. The estimation of clothing level is another challenging task which increases system cost for additional sensing devices such as a depth camera [111].

One of our future challenges is to construct a personal model to improve estimation performance. Laftchiev et al. [112] proposed a weakly-supervised algorithm to learn the individual thermal comfort preferences based on an autoencoding framework. Our method may work effectively with such a model which can learn personal preference. He et al. [113] proposed the use of mobile device batteries as thermometers, which may be helpful in measuring the air temperature around the subject for more accurate TSV estimation. Because we collected the TSV reports using a smartphone application, the smartphone thermometer may be useful to collect personal training data without any additional sensors.

4.6 Conclusion

In this chapter, we firstly investigated the effectiveness of combining a thermal camera and a wristband sensor for continuous and accurate thermal comfort estimation. Through the preliminary study using real data, we confirmed that our method is effective to detect the redundant use of air conditioning. Secondly, we proposed TSVNet, which is based on deep learning, for TSV estimation. TSVNet combines time-series data with opportunistic data to address a dynamic environment. We demonstrated that the estimation performance improved in the dynamic environment by utilizing time-series data measured by a wristband sensor. In addition, the thermal sensation measured in a real-world setting is imbalanced with a neutral bias. To address this problem, we also proposed the data balancing method for datasets that contain time-series data. The result indicated that the balancing method effectively extracted features. We also investigated the duration of time before the measurement that affects the current thermal sensation. The result showed that the previous 10 minutes is sufficient for the estimation.

Chapter 5

Physiological Data Balancing

5.1 Introduction

Machine learning is one of the most commonly used approaches for a wide variety of applications, including healthcare. Many of the healthcare applications typically use time-series data such as heart rate and body temperature [114, 115] to estimate psychological or physiological states of humans using machine learning algorithms. For training estimators, data collection through real experiments is essential. The challenge in data collection, especially in healthcare applications, is data imbalance, i.e., the distribution of the dataset is non-uniform. This is natural because minor cases do not frequently happen in the real world. Especially in thermal comfort estimation, various machine learning-based estimation methods have been proposed by using physiological and environmental data collected in real environment. Datasets used in such works are mostly imbalanced because hot/cold environments rarely appear in an actual air-conditioned environment. The imbalanced dataset causes a problem that the estimator tends to focus more on the normal samples, neglecting the rare samples. Researchers have applied data balancing by augmenting the rare samples to mitigate the imbalances of such thermal comfort datasets. The imbalanced data causes classifiers to be naturally biased towards the majority class, leading to the performance degradation for important and interest minority samples [116, 17]. This is because standard machine learning methods usually seek the minimization of training errors. We may be able to collect data in minor cases by designing experiment protocols carefully for some researches. Nevertheless, such data collection is limited in controlled environments. Therefore, it is difficult to collect data in minor cases in uncontrolled environments for researches in the healthcare domain.

In order to deal with the imbalanced datasets, several strategies have been developed. A data preprocessing-based approach called data balancing [117] is one of the solutions. The data balancing changes the data distributions to make the estimators focus on the samples, which are important for the users. The data balancing is also a way of pre-processing the data to increase its size, diversity, and robustness of models. Oversampling and undersampling are used for balancing the dataset by

augmenting minor data and discarding major data, respectively. The former approach is applied based on existing algorithms such as Synthetic Minority Oversampling Technique (SMOTE) [17]. SMOTE augments the data in minor classes by interpolation for classification problems. We can apply them to any existing machine learning algorithms because they adjusts the data distributions before machine learning.

Chawla et al. [17] proposed SMOTE, which is a predominant data augmentation for data balancing. They showed the advantages of this approach compared to other alternative sampling techniques on several real-world problems by using several classification algorithms. It generates new minor class samples by linear interpolation between neighboring original samples. However, in many real-world applications, we have high-dimensional data that includes numerical and categorical variables, which we cannot generate with a simple interpolation-based augmentation. Another approach for data augmentation is known as generative models [76] based on deep learning such as Generative Adversarial Networks (GANs) [77]. This is a learning-based approach which generates synthetic data using a pre-trained generative model from real data. Conditional GANs (cGANs) [118] is a variant of GANs, which generates data with a class label. For balancing a thermal comfort dataset, Quintana et al. proposed GANs-based data balancing [119]. This method utilized conditional Wasserstein GAN-gradient penalty (comfortGAN) to address the problem of the imbalanced thermal comfort dataset. They showed GAN-based data balancing could augment the thermal comfort dataset to improve the performance of an estimator. However, the class imbalance may cause biased generation for GANs because balancing methods using GANs usually train their generators and discriminators from the imbalanced datasets.

In the former part in this chapter, we propose a data balancing method for thermal comfort datasets using comfortGAN with a weighted loss function to address this problem. The key idea is introducing a weighted loss function based on the number of samples in each class while training comfortGAN. We evaluate our method by using four thermal comfort datasets including three open datasets used in Ref. [119]. We also use three evaluation metrics propoosed in Ref. [119], which are variability, diversity, and accuracy to evaluate the performance of data balancing. The result shows that our approach helps to improve the estimation of the thermal comfort, especially for a small dataset. Our contribution is that we introduce the loss function considering the difference among the numbers of samples in all classes to avoid biased training towards major classes.

In the field of physiological sensing, estimations of human states such as thermal sensation using machine learning often regard the target problems as classification problems for simplicity [16, 59]. However, to achieve estimations of human states with finer granularity, regression is more appropriate than classification because regression estimates numerical values. Because regression problems output continuous numerical values, data balancing algorithms for classification problems are not suitable for regression problems. For data balancing in regression problems, some algorithms such as SMOTER [18] were proposed based on SMOTE. SMOTER divides the distribution of the numerical target values

Table 5.1: Summary of balancing methods

Method	time-series	target problem
SMOTE	No	classification
TS_SMOTE	Yes	classification
SMOTER	No	regression
Proposed method	Yes	regression

into the major and minor values based on the relevance score of the target value. After that, it applies undersampling for major cases and oversampling for minor cases. The oversampling strategy is designed for numerical values based on SMOTE. It uses the weighted average between the target values of the two minor cases based on the distance between them. Through the above steps, the dataset for a regression problem is balanced. However, SMOTER does not consider time-series feature values commonly used in psychological or physiological state estimation.

In this chapter, we secondly propose a data balancing method for regression with time-series feature values based on SMOTER. In the proposed method, we generate relevance function using probabilistic density function generated by kernel density estimation. This method does not assume the distribution of the response variable. Therefore, it does not require the parameter tuning by the user. To consider the temporal dependency of time-series data, we extend a distance function. To define the distance between time-series samples, we use Dynamic Time Warping (DTW) distance as used in TS_SMOTE [22]. TS_SMOTE is designed to extend SMOTE to time-series data. Our method interpolates synthetic time-series using the weighted average and the DTW distance. Table 5.1 summarizes the difference between our method and other balancing methods in terms of the capability to deal with time-series features and target problem types. As far as we know, our method is the first to achieve data balancing for regression problems with time-series feature values.

For evaluation, we apply our method to two imbalanced datasets with time-series feature values. The first dataset is the thermal sensation dataset, which consists of time-series physiological data measured by a wristband sensor as a feature value and thermal sensation vote (TSV) as a target value. The second dataset is the core body temperature dataset. The core body temperature is measured by a tympanic temperature sensor during exercise. The feature values are measured by a wristband sensor, a chest strap sensor, and an environmental sensor. The results show that our time-series augmentation improves the performance of regression models for minor cases with a little decline in the mean average error.

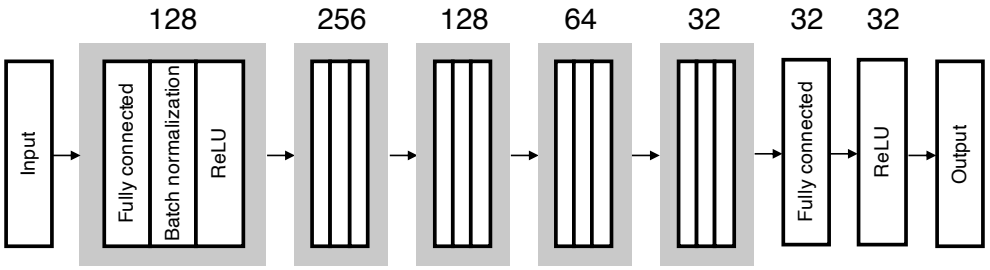


Figure 5.1: Architecture of generator

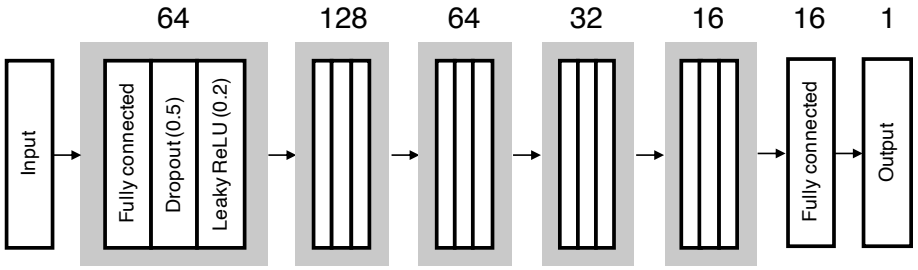


Figure 5.2: Architecture of discriminator

5.2 Proposed Method

5.2.1 Weighted comfortGAN

Weighted comfortGAN uses the original comfortGAN architecture to augment the thermal comfort dataset. The architecture is based on cGANs, which is a variant of GANs. The generation and discrimination of samples by cGANs are conditioned on the class label. The generator and discriminator architectures are summarized in Figure 5.1 and Figure 5.2, respectively.

For the generator, we use five blocks. This block consists of three layers, i.e., a fully connected layer, a batch normalization layer, and a Rectified Linear Units (ReLU) layer. The numbers of nodes per layer are 128, 256, 128, 64, and 32 from the first to the last layer. After these blocks, a fully connected layer with ReLU is used to generate the data. For the discriminator, we use five blocks. This block consists of three layers, i.e., a fully connected layer, a dropout layer whose dropout rate is 0.5, and a leaky ReLU layer with a slope of 0.2. After these blocks, a fully connected layer is used to output the label, i.e., real or fake.

The loss function of the original comfortGAN is based on the idea of Wasserstein GAN-gradient penalty (WGAN-GP). WGAN-GP was proposed to facilitate training convergence and stability compared to GANs. We modify the loss function of the discriminator (L_d) with the weight considering

the number of samples in each class as follows.

$$L_d = \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g}[D(\tilde{\boldsymbol{x}})W(\boldsymbol{x})] - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_r}[D(\boldsymbol{x})W(\boldsymbol{x})] \quad + \lambda \mathop{\mathbb{E}}_{\hat{\boldsymbol{x}} \sim \mathbb{P}_{\hat{\boldsymbol{x}}}}[(\parallel \nabla D(\hat{\boldsymbol{x}})W(\boldsymbol{x}) \parallel_2 - 1)^2], \tag{5.1}$$

where $\hat{\boldsymbol{x}} = \epsilon \boldsymbol{x} + (1-\epsilon)\tilde{\boldsymbol{x}}, \epsilon \sim U[0,1]$. $D(\boldsymbol{x})$ is the output of the discriminator for input of $\boldsymbol{x}$. The real sample $\boldsymbol{x}$ and the fake sample $\tilde{\boldsymbol{x}}$ by the generator is input to the discriminator.

Also, we define the loss function of the generator (L_g) using the weight as follows.

$$L_g = -\mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g}[D(\tilde{\boldsymbol{x}})W(\boldsymbol{x})]. \tag{5.2}$$

The weight function $W(\boldsymbol{x})$ is based on the inverse ratio of the number of samples in each class in the training data as follows.

$$W(\boldsymbol{x}) = \frac{|samples|}{N(\boldsymbol{x})|classes|}, \tag{5.3}$$

where $N(\boldsymbol{x})$ is the number of samples in class $c(\boldsymbol{x} \in c)$. $|samples|$ is the total number of samples in the training data, and $|classes|$ is the number of classes. In order to keep the expected value of the loss the same as the original method, we determine the size of the weight. For example, weights for three classes, i.e., warm, neutral, and cool, with the numbers of samples are 100, 400, and 100, are determined as 2.0, 0.5, and 2.0, respectively.

5.2.2 Problem Definition of Time-series Augmentation for Regression

Imbalanced regression is a sub-class of regression problems [120]. In this setting, given a training set $\mathcal{D} = \{\langle \boldsymbol{x}_i, y_i \rangle\}_{i=1}^N$, the goal is to obtain a model $m(\boldsymbol{x})$ that approximates an unknown regression function $Y = f(\boldsymbol{x})$ as defined in Ref. [120]. In the problem we address, $\boldsymbol{x}$ is a feature vector that contains time-series data from N_f sensors as follows.

$$\boldsymbol{x}_i = \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{N_f}\} \tag{5.4}$$

$$\boldsymbol{f}_j = \{s_1, \ldots, s_{N_s}\} \tag{5.5}$$

where N_f is the number of time-series features, and T is the length of the sample of the time-series $\boldsymbol{f}_j$. In this paper, N_s is the same constant value for any $\boldsymbol{f}_j$. This is because basic learning approaches for time-series, such as Recurrent Neural Network (RNN), assume the same length of time-series as their input.

5.2.3 Time-series Augmentation for Regression

We propose a data augmentation method based on the existing method proposed in Ref [121]. The overview of the existing data balancing method proposed in Ref [121] is shown in Figure 5.3. The input and output are an imbalanced original dataset with time-series features and a balanced dataset, respectively. First, a relevance function $\phi(y)$ is generated based on the distribution of the target value y in the original dataset as proposed in SMOTER. The relevance function is automatically generated

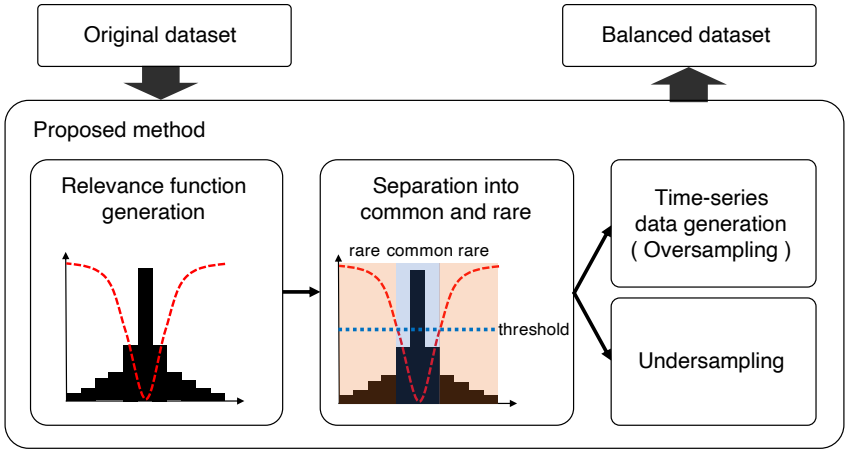


Figure 5.3: Overview of the existing data balancing method

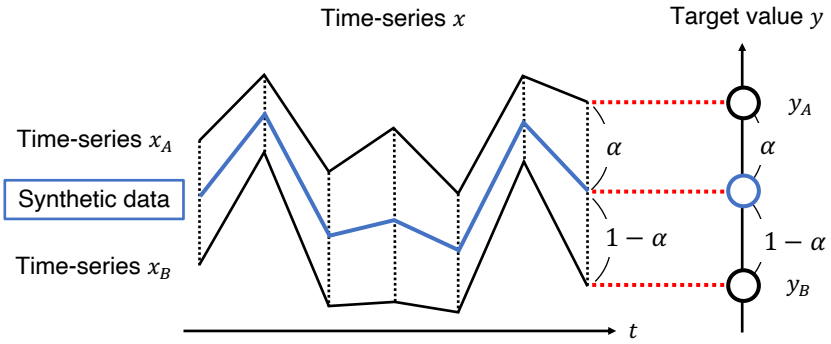


Figure 5.4: Key idea of time-series data generation for regression

based on a probability density function (PDF) [122]. Second, the original distribution is separated into minor data D_r and major data D_c by a user-defined threshold t_E as follows.

$$D_r = \{\langle \boldsymbol{x}, y\rangle \in D | \phi(y) > t_E\} \tag{5.6}$$

$$D_c = D \setminus D_r \tag{5.7}$$

After this separation, oversampling and undersampling are carried out on D_r and D_c, respectively.

Based on D_r, we apply time-series data generation. Figure 5.4 illustrates our key idea based on SMOTER. A synthetic data is interpolated between two samples in D_r with a ratio α, which is determined by the weighted average of them. The data generation in SMOTER needs to define a distance of a pair of $\boldsymbol{x}$. Our main contribution is combining the DTW distance with SMOTER. The distance can be used for time-series interpolative generation conserving original time-series features, such as shape [22]. As shown in Figure 5.4, synthetic points in the generated time-series are interpolated

between a pair, which is given by DTW, using the ratio α. This existing method succeeded to generate time-series data conserving the trend of the original data.

However, the generated data does not have natural distribution of the response variable y, i.e., thermal sensation. This is because the existing method generate the data based on the threshold between the common samples and the rare samples. This threshold causes generating the imbalanced distribution of the response variable y. Therefore, we propose the time-series augmentation method with probabilistic extraction from rare samples. The proposed method extract a rare sample which is selected as an original to mimic a synthetic sample, based on the probability calculated the relevance function. In this extraction procedure, the probability to select a rarer sample is higher. In addition, we propose the generate method of relevance function. The proposed method is a nonparametric method that does not require parameter tuning for users without expertise in machine learning by not assuming the distribution of the response variable y. It is based on the idea that the PDF can be estimated without requiring parameters by assuming the a distribution for the kernel density estimation (KDE) [123, 124], which is a PDF estimation method. It uses Scott's rule whose fitness to real-world data is empirically guaranteed. The PDF f_{dens} is calculated from KDE as following formula.

$$f_{dens}(x) = \frac{1}{nh} \sum_{i=0}^{n} K\left(\frac{x - x_i}{h}\right), \tag{5.8}$$

where n is the number of samples in the training data, h is the band width, and K is the kernel function. In the proposed method, we use following Gaussian function as the kernel function.

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{(-x^2/2)}. \tag{5.9}$$

The band width h is calculated from Scott's rule [125], which do not need given parameters as follows.

$$h \approx 1.06 u n^{(-1/5)}, \tag{5.10}$$

where u is the sample standard deviation of the response variable in the training data. The relevance function is determined based on the distribution of the response variable calculated from KDE. The generation procedure is summarized in Figure 5.5. The method for generating the relevance function f_r in the proposed method is shown in Algorithm 3. By standardizing the PDF f_{dens} obtained by KDE, a function f_{norm} is generated with a maximum value of 1 and a minimum value of 0. By subtracting the function f_{norm} from one, we generate the relevance function f_r, which returns values closer to 1 for rare sample and closer to 0 for common sample. After the above procedures, we extract samples based on the probability using the relevance function f_r. The pseudocode of the main algorithm for data augmentation is shown in Algorithm 4. We use the same procedure to generate time-series sample, i.e., Algorithm 2 after the extraction of the sample. The proposed method is summarized in Figure 5.6.

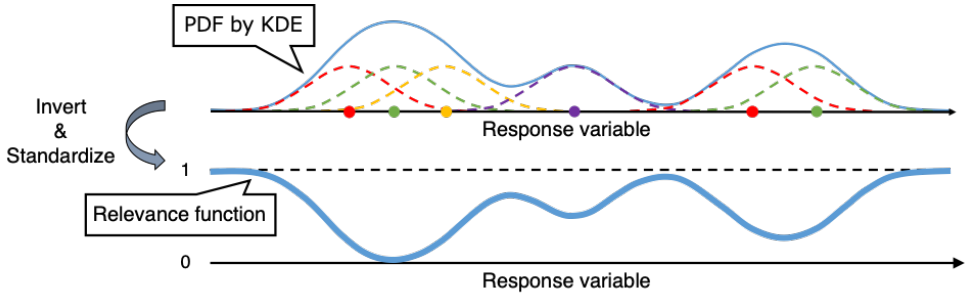


Figure 5.5: Generation procedure of relevance function

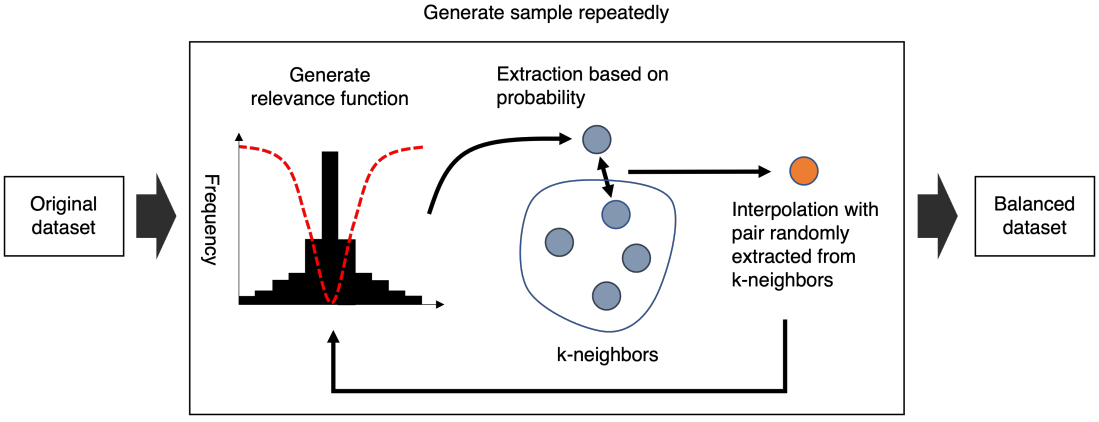


Figure 5.6: Data generation procedure in proposed method

Algorithm 3 Relevance function generation

Input: $\boldsymbol{y}$ - The list of the response variable
Output: f_r - Relevance function
  # Kernel density estimation with Scott's rule
  $f_{dens} \Leftarrow$ kernelDensityEstimation($\boldsymbol{y}$)
  # Normalize density function
  $f_{norm} \Leftarrow f_{dens}/\max(f_{dens})$
  # Generate relevance function as complementary standardized density function
  $f_r \Leftarrow 1 - f_{norm}$
  return f_r

5.3 Evaluation on Weighted comfortGAN

5.3.1 Dataset

We evaluate our method by using four datasets, including three open datasets used in Ref. [119]. The first dataset was collected in Ref. [126] from 77 participants in Pittsburgh. The experiment was conducted for three hours in the controlled environment for one year. They collected environmental and physiological data with thermal comfort subjective feedback, i.e., uncomfortably cold, slightly uncomfortably cold, comfortable, slightly uncomfortably warm, and uncomfortably warm, via a mobile application in the environment where the room temperature is fixed. The second dataset was collected in Ref. [127] from 30 participants in an educational building in Singapore over two weeks. The subjects reported their thermal preference in three classes, i.e., prefer warmer, comfortable, and prefer cooler. The third dataset was collected in Ref. [128]. This dataset is from an open-source research database, which was collected in the two decades of thermal comfort field studies around the world. The subjects reported their thermal sensation in the American Society of Heating, Refrigerating, and Air-Conditioning Engineers' (ASHRAE) 7-point thermal scale [97]. The fourth dataset is used by TSVNet [121]. This dataset was collected for a total of 123 days from 21 subjects in static and dynamic environments. The subjects also reported their thermal sensation by ASHRAE 7-point thermal scale. Although opportunistic and time-series features are used in TSVNet, we use only the opportunistic features because we cannot directly input the time-series features to the proposed method. In addition, we merged the hot/cold samples into the warm/cool samples because of the lack of the rare samples (i.e. hot and cold). This is because there may be a case that the rare samples of the thermal sensation are not included in the training data. The summary of the datasets is shown in Table 5.2.

5.3.2 Evaluation Metric

We evaluate our method with three metrics for data balancing proposed in Ref. [119]: (1) variability of generated samples, (2) diversity of generated samples with respect to the training set, and (3) machine learning efficacy or quality of generated samples on classification tasks. The variability of generated samples is defined as the difference between generated samples randomly drawn from the

Algorithm 4 Generating augmented dataset

Input: $\mathcal{D}$ - A dataset
$\%o$ - Percentages of oversampling
k - Number of neighbors used in case generation
Output: $\mathcal{D}_{gen}$ - Generated new cases

$newCases \Leftarrow \{\}$
$ng \Leftarrow \%o * size(\mathcal{D})/100$
for $i \Leftarrow 1$ **to** ng **do**
  $f_r \Leftarrow$ genRelevanceFunction$(\mathcal{D} \cup newCases)$
  $case \Leftarrow$ extractCase$(\mathcal{D}, f_r)$
  $nns \Leftarrow \text{KNN}(k, case, \mathcal{D} \setminus \{case\})$
  $n \Leftarrow$ randomly choose one of the nns
  $\alpha \Leftarrow$ randomly choose in $[0, 1]$
  $new[y] \Leftarrow min(case[y], n[y]) + \alpha|case[y], n[y]|$
  for all $f \in$ features **do**
    $pairs \Leftarrow \text{DTW}(case[f], n[f])$
    for $t \Leftarrow 1$ **to** $|pairs|$ **do**
      $(t_{new}, v_{new}) \Leftarrow$ timePoint$(pairs[t]_{case}, pairs[t]_n, case[f], n[f], \alpha)$
      $new[f, t_{new}] \Leftarrow v_{new}$
    end for
  end for
  $newCases \Leftarrow newCases \cup \{new\}$
end for
return $newCases$
function timePoint$(t_a, t_b, ts_a, ts_b, \alpha)$
  $t_{new} \Leftarrow min(t_a, t_b) + (t_a + t_b)/2$
  $v_{new} \Leftarrow min(ts_a[t_a], ts_b[t_b]) + \alpha|ts_a[t_a] - ts_b[t_b]|$
  return t_{new}, v_{new}
end function

generated dataset. The diversity of generated samples with respect to the training set is defined as the difference between closest pairs from generated and training set calculated as Euclidean distance. The machine learning efficacy or quality of generated samples on classification tasks is defined as classification accuracy on a balanced mixture of generated and training samples datasets. It reflects how the balancing method improves the estimator's performance. In addition to them, we evaluate our method with metrics generally used for classification tasks: (4) precision, (5) recall, and (6) F1-score. They can evaluate how the classification is balanced because they are the averages of each metric in each class. In addition to these metrics, we conducted further evaluation on the similarity of the distributions between the original dataset and the generated dataset. We selected the air temperature as the index of the similarity because the air temperature is the most effective feature for the thermal comfort estimator [20]. As used in Ref. [119], we train random forest classifier for the thermal comfort estimation.

We compare the proposed method with three methods. The first one is a baseline method, which estimates thermal sensation without any data balancing. The second one is a SMOTE-based data

Table 5.2: Summary of datasets

Dataset	# of samples	# of features	# of classes
Ref. [126]	2067	9	5
Ref. [127]	1474	10	3
Ref. [128]	66397	6	7
Ref. [121]	1688	14	5

Table 5.3: Result of variability

Model	Variability			
	[126]	[127]	[128]	[121]
Baseline	25.15	146.25	18.39	24.65
SMOTE	48.36	**167.29**	17.89	22.66
comfortGAN	**67.92**	137.23	**29.16**	**27.70**
Proposed	61.72	139.55	15.86	19.51

Table 5.4: Result of diversity

Model	Diversity			
	[126]	[127]	[128]	[121]
Baseline	3.45	5.14	0.49	3.34
SMOTE	3.13	5.90	0.61	3.91
comfortGAN	229.05	90.13	15.63	109.95
Proposed	**232.83**	**90.53**	**17.72**	**111.13**

balancing applied before the training of the estimator. The third one is comfortGAN, a state-of-the-art method for learning-based data balancing to classify thermal comfort. After these pre-processing, we randomly chose 70% samples for training of the random forest and the remaining 30% for the test.

5.3.3 Result

The results of the six metrics are summarized in Table 5.3, Table 5.4, Table 5.5, Table 5.6, Table 5.7, and Table 5.8. As shown in Table 5.4 and Table 5.5 diversity and accuracy were improved or same compared to comfortGAN. Especially, the proposed method increase diversity for all datasets compare to comfortGAN. This result shows that the weighted loss function works to generate different samples from the original dataset while conserving the original trend. This is our contribution, which has the possibility to augment the sensing data and estimate how the human body responds in a virtual situation, i.e., simulation between a building environment and occupants. However, variability decreased in the three datasets compared to comfortGAN because the proposed method is more converged throughout the balanced training as shown in Table 5.3. This variability reduction means the generated dataset does not have a wider distribution than the dataset generated by comfortGAN.

Table 5.5: Result of accuracy

Model	Accuracy			
	[126]	[127]	[128]	[121]
Baseline	0.68	0.69	**0.42**	0.58
SMOTE	0.68	0.65	0.37	0.58
comfortGAN	0.67	0.69	**0.42**	**0.63**
Proposed	**0.69**	**0.70**	**0.42**	**0.63**

Table 5.6: Result of macro-averaged precision

Model	Precision			
	[126]	[127]	[128]	[121]
Baseline	**0.60**	**0.65**	**0.35**	0.43
SMOTE	0.57	0.59	0.28	0.43
comfortGAN	0.58	**0.65**	**0.35**	0.45
Proposed	0.59	**0.65**	**0.35**	**0.48**

Table 5.7: Result of macro-averaged recall

Model	Recall			
	[126]	[127]	[128]	[121]
Baseline	0.55	0.55	0.27	**0.46**
SMOTE	**0.57**	**0.59**	**0.30**	0.45
comfortGAN	0.55	0.55	0.27	0.38
Proposed	0.56	0.56	0.27	0.40

However, there is a trade-off between the wide range of distribution of the training data and the accuracy of the estimator because completely different samples can disturb the convergence of the training. Therefore, the generation of similar distribution is important for the convergence of the training.

In order to evaluate the similarity between the balanced dataset and the original dataset, we show the mean and median of air temperature in each class in the original dataset, the balanced dataset by comfortGAN, and the balanced dataset by the proposed method in Table 5.9 and Table 5.10. Both methods were trained using the dataset of Ref. [121] in the table. We note that the neutral samples after data balancing are same as the samples before the data balancing because new neutral samples are not generated due to the majority. This result indicates that the proposed method can generate the data conserving the distribution of the original dataset. This is because the training of the generator may have been well converged owing to the weight function to focus on the rare samples. Therefore, because of the convergence of the training, the variability slightly drops compared to comfortGAN. For further evaluation, the distributions of the air temperature in each class generated by comfortGAN and the proposed method are shown in Figure 5.7. Both methods were also trained using the dataset of Ref. [121]. This result shows that the proposed method can generate air temperature data whose

Table 5.8: Result of macro-averaged F1-score

Model	F1-score			
	[126]	[127]	[128]	[121]
Baseline	0.56	0.58	**0.29**	**0.44**
SMOTE	**0.57**	**0.59**	**0.29**	**0.44**
comfortGAN	0.56	0.58	0.28	0.40
Proposed	**0.57**	**0.59**	**0.29**	0.41

Table 5.9: Mean of air temperature distribution in each class in the datasets

Thermal sensation	Mean of air temperature [°C]		
	Original dataset	comfortGAN	Proposed
Warm	27.93	26.02 (−1.91)	27.09 (**−0.82**)
Slightly warm	26.46	25.35 (−1.11)	26.69 (**+0.23**)
Neutral	25.58	25.58 (±0)	25.58 (±0)
Slightly cool	23.80	23.37 (−0.43)	24.14 (**+0.34**)
Cool	22.14	23.26 (+1.12)	23.13 (**+0.99**)

Table 5.10: Median of air temperature distribution in each class in the datasets

Thermal sensation	Median of air temperature [°C]		
	Original dataset	comfortGAN	Proposed
Warm	27.80	25.92 (−1.88)	27.08 (**−0.72**)
Slightly warm	26.00	25.13 (−0.87)	26.70 (**+0.70**)
Neutral	25.20	25.20 (±0)	25.20 (±0)
Slightly cool	24.05	23.39 (−0.66)	24.15 (**+0.10**)
Cool	21.70	23.11 (**+1.41**)	23.14 (+1.43)

distribution is more similar to the original dataset than comfortGAN. The result for all datasets had similar trends, i.e., the proposed method generated more similar to the original dataset than comfortGAN. However, we could not show all the results due to the limitation of the number of pages.

Table 5.6, Table 5.7, and Table 5.8 show the metrics for evaluation of classification performance for the entire dataset. This result shows that the proposed method improves these metrics without any degradation in the performance for all datasets. This result is important for the evaluation of the performance of the classifier trained by the imbalanced dataset because these metrics reflect the imbalance of the classification performance for both of the classes, i.e. neutral and extreme sensations. This improvement shows the proposed method can generate rare samples conserving the distribution of the original dataset for thermal comfort data balancing. There are some cases where the baseline or SMOTE shows the best score, such as F1-score for the dataset of Ref. [121]. In this case, their accuracy is lower than the proposed method. This result means there is a trade-off between the accuracy and the macro-averaged F1-score. In terms of the trade-off, the proposed method achieves

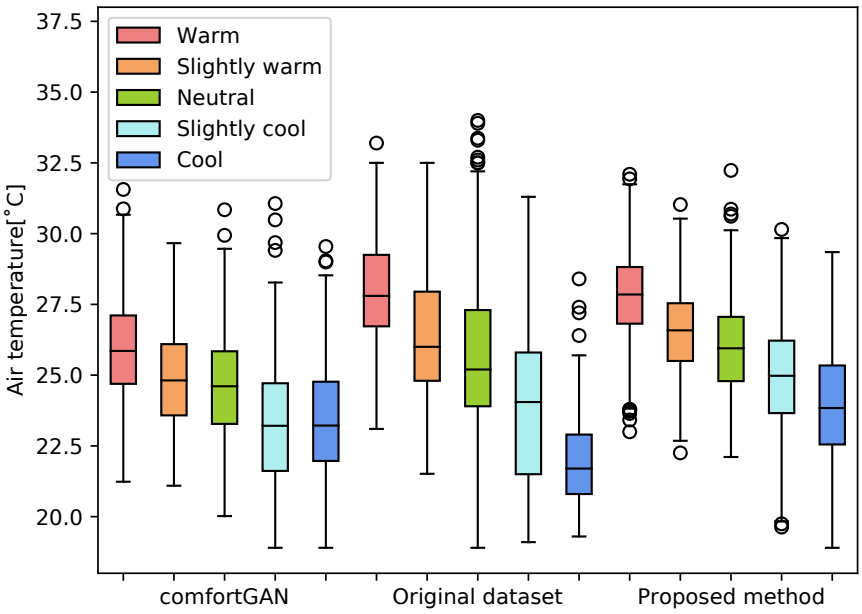


Figure 5.7: Distributions of air temperature in the datasets

better performance than the baseline. This is because the proposed method can generate data with variability and diversity similar to the original dataset. This leads to the slight improvement of the performance of the thermal comfort classification by the proposed method.

5.4 Evaluation on Time-series Augmentation for Regression

5.4.1 Dataset

The first dataset is a thermal sensation dataset collected from 21 subjects. The target value of the dataset is Thermal Sensation Vote (TSV). The subjects reported TSV within the range of [-3.5, 3.5] by moving the seek bar up and down on our smartphone application. The scale is called the American Society of Heating, Refrigerating, and Air-Conditioning Engineers' seven-point thermal scale [97], which is widely used as the metrics of human thermal sensation; the seven levels range from -3 to +3 (Cold, Cool, Slightly cool, Neutral, Slightly warm, Warm, Hot). In total, we collected 1686 TSV inputs. The distribution of the TSV inputs is shown in Figure 5.8. Red hatched areas in the figure highlight the minor distribution. Because most of the data are collected in an air-conditioned environment, most of the TSVs labeled by the subjects are +1 (Slightly warm), 0 (Neutral), or -1 (Slightly cool).

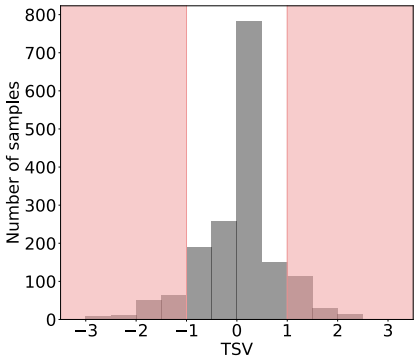


Figure 5.8: Distribution of TSV

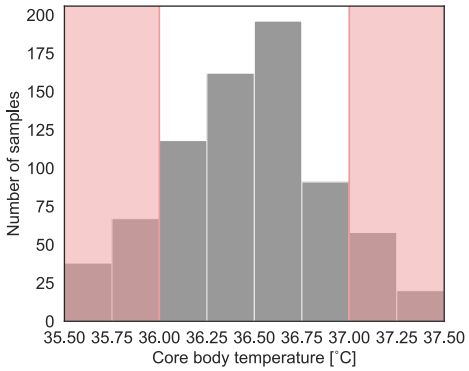


Figure 5.9: Distribution of core body temperature

Three feature values are heart rate, skin temperature, and electrodermal activity collected by an E4 wristband sensor [10]. They are measured continuously as time-series data.

The second dataset is a core body temperature dataset collected from 13 subjects while exercising. The target value is the core body temperature recorded by a cosinuss° C-Temp [129], which can measure tympanic temperature. The feature values are measured by two physiological sensors and an environmental sensor. The first one of the physiological sensors is WHS-3 [130], which is a wearable heart rate sensor with a chest strap. We use heart rate and in-cloth temperature from WHS-3 as feature values. The second one is the E4 wristband sensor as used in the TSV dataset. In addition to the three feature values used in the TSV dataset, we also use acceleration from it. Also, we use an environmental sensor to measure air temperature and relative humidity as time-series. In total, eight features are input to an estimator. We collected 750 pairs of the core body temperature and ten minutes of the eight time-series data through the experiment. The distribution of the core body temperature is shown in Figure 5.9.

5.4.2 Evaluation Metric

We evaluate the proposed method with evaluation metrics based on precision P_r, recall R_r, and F-measure F_r for regression problems as proposed in Ref. [131]. Intuitively, the metrics become larger (i.e. better) when an estimator outputs closer values in rarer cases. The definition is given below.

$$P_r = \frac{\sum_{\phi(\hat{y}_i)>t_E} \phi(\hat{y}_i)\alpha(\hat{y}_i, y_i)}{\sum_{\phi(\hat{y}_i)>t_E} \phi(\hat{y})}, \tag{5.11}$$

$$R_r = \frac{\sum_{\phi(y_i)>t_E} \phi(y_i)\alpha(\hat{y}_i, y_i)}{\sum_{\phi(y_i)>t_E} \phi(y_i)}, \tag{5.12}$$

$$F_r = \frac{2P_rR_r}{P_r + R_r}, \tag{5.13}$$

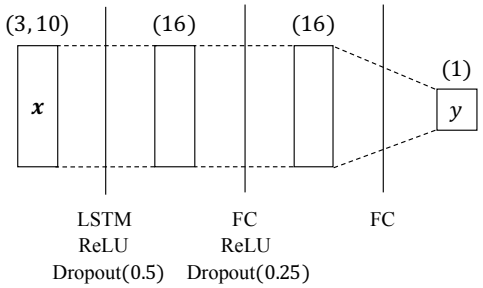


Figure 5.10: TSV estimator

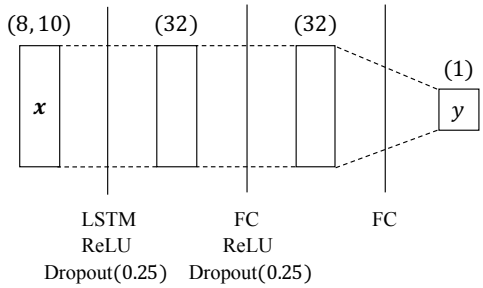


Figure 5.11: Core body temperature estimator

where y_i and $\hat{y_i}$ are an actual value and an estimated value by inputting x_i to the estimator, respectively. As defined in Ref. [131], the function α is defined as follows.

$$\alpha(\hat{y_i}, y_i) = I(L(\hat{y_i}, y_i) \leq t_L)(1 - \exp(\frac{-k(L(\hat{y_i}, y_i) - t_L)^2}{{t_L}^2})), \tag{5.14}$$

where I is an indicator function which is one if its argument is true and zero otherwise. t_L is a threshold defining an acceptable error within the domain a metric loss function L, e.g., the absolute deviation. k is a positive number that determines the shape of the function.

5.4.3 Estimator Design

To estimate the numerical target value y, we construct a deep learning-based estimator for each dataset using an LSTM layer. Figures 5.10 and 5.11 show estimators for the thermal sensation dataset and core body temperature dataset, respectively. The shape of $\boldsymbol{x}$ is (N_f, N_s) as defined in Section 5.2.2. The input $\boldsymbol{x}$ consists of three time-series data whose length N_s is ten.

5.4.4 Generation of Relevance Function

The result of the generation of relevance function is shown in Figure 5.12, Figure 5.13, Figure 5.14, and Figure 5.15. The relevance function generated using the existing method is shown in Figure 5.12 and Figure 5.12. As shown in these figures, the distributions of the response variable y are biased. This is because the existing method generate samples by separation into rare class and common class. As a result, the "common" case in the rare class is selected as the original sample for the generation with higher probability. The dataset generated from the existing method probably causes biased estimation. On the other hand, Figure 5.13 and Figure 5.13 show the result of the generation of relevance function by the proposed method. These samples are generated until the size of the augmented dataset becomes four times as large as the size of the original dataset. As shown in these figures, the relevance function generated by the proposed method is not biased compare to the relevance function generated by the existing method.

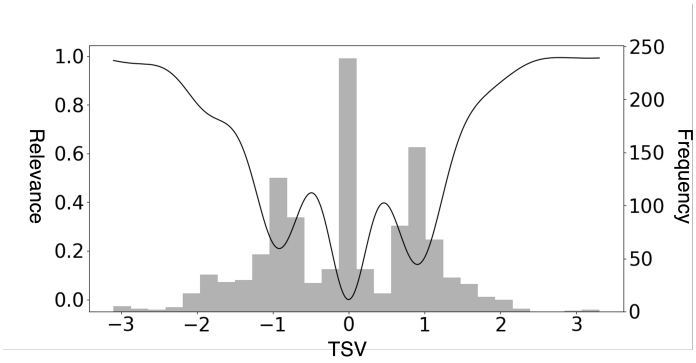


Figure 5.12: Relevance function generated by existing method for TSV dataset

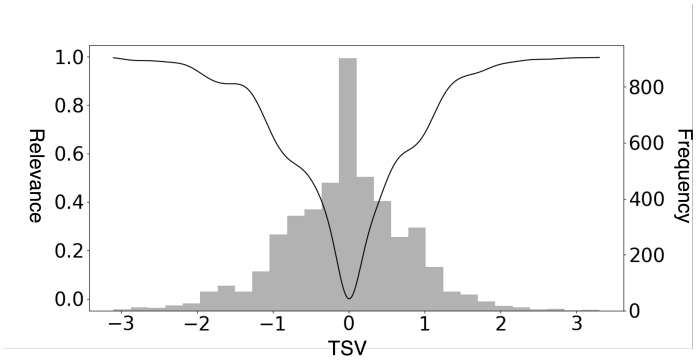


Figure 5.13: Relevance function generated by proposed method for TSV dataset

5.4.5 Effect of Data Balancing

We evaluate the proposed method through comparison with baseline, which is without any data balancing, and SMOTER-based time-series interpolation with Euclidean distance. The latter method generates a time-series sample by interpolating a synthetic sample between corresponding samples in original time-series data as shown in Figure 5.16. The figure shows an example of the time-series generation by each method given two time-series samples. As illustrated in the figure, SMOTER fails to inherit features of the original time-series samples such as the maximum and minimum values. On the other hand, the proposed method generates a synthetic time-series sample with the features of the original time-series samples such as the shape, maximum, and minimum values.

Table 5.11 and Table 5.12 show the evaluation results of the TSV dataset and the core temperature dataset, respectively. The methods x2, x3, and x4 means the number of samples generated by the augmentation method. For example, x2 means that the number of samples in the augmented dataset is twice as many as the number of samples in the original dataset. We note that in addition to the evaluation metrics defined in Section 5.4.2, we evaluate the mean absolute error (MAE). Generally,

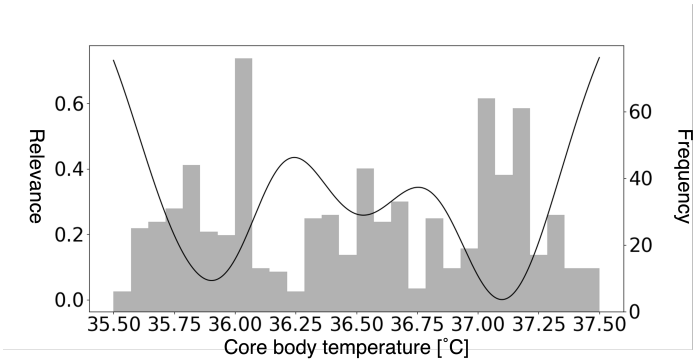


Figure 5.14: Relevance function generated by existing method for core body temperature dataset

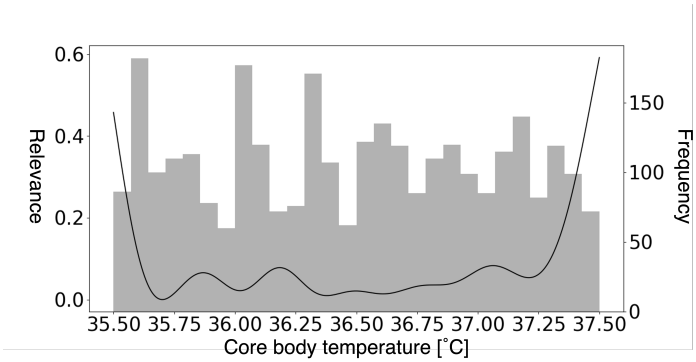


Figure 5.15: Relevance function generated by proposed method for core body temperature dataset

MAE becomes low for the test data is biased when the estimator is also biased. Therefore, there is a trade-off between F_r and MAE when we evaluate the estimator using the imbalanced dataset. We set the parameters for Table 5.11 as $t_E = 0.5$ and $k = 5$ for the existing method, and $t_L = 1$ for the calculation of F_r. For Table 5.12, we set them as $t_E = 0.5$, $k = 5$, and $t_L = 1$. As shown in the table, the baseline, which is an estimator trained using imbalanced dataset, result in low F_r for rare cases. This result is similar to the imbalanced classification problem. The result of F_r shows that the baseline method fails to estimate rare cases within the acceptable error. In addition, the result of SMOTER with Euclidean distance showed lower improvement than the proposed method in Table 5.11. This is because the time-series generation based on the DTW distance can generate time-series samples close to the original time-series samples. As a result, the proposed method remarkably enhances the estimator's performance for rare cases with a little decline of the MAE compared with the baseline. In Table 5.12, the result of the proposed method shows good trade-off compared with the result of the SMOTER. The proposed method's MAE is improved from the baseline because the estimator can

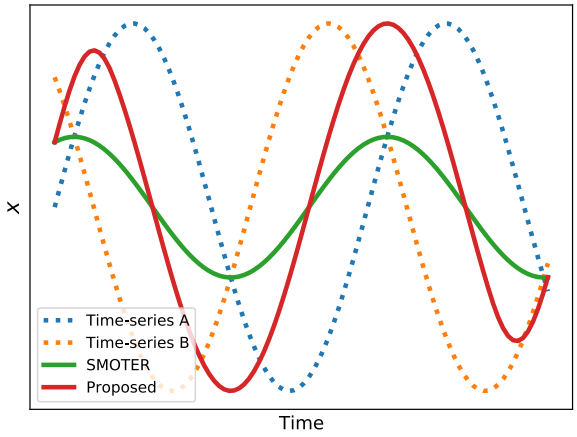


Figure 5.16: Example of time-series generation by proposed method

Table 5.11: Estimation result for TSV dataset

Method	Baseline	SMOTER (Euclidean)				Proposed (DTW)			
		Existing	x2	x3	x4	Existing	x2	x3	x4
F_r	0.00	0.00	0.11	0.14	0.25	0.00	0.12	0.25	0.25
MAE	0.49	0.50	0.59	0.69	0.72	0.49	0.58	0.66	0.71

Table 5.12: Estimation result for core temperature dataset

Method	Baseline	SMOTER (Euclidean)				Proposed (DTW)			
		Existing	x2	x3	x4	Existing	x2	x3	x4
F_r	0.00	0.57	0.57	0.53	0.53	0.57	0.58	0.61	0.62
MAE	0.42	0.50	0.47	0.47	0.47	0.49	0.45	0.44	0.46

estimate in a larger range than the baseline, which is overfitted to a common sample.

5.5 Conclusion

In this chapter, firstly, we proposed a data balancing method for thermal comfort datasets using conditional Wasserstein GAN with a weighted loss function. We evaluated the performance of our data balancing method using six metrics used in the existing work. The result shows our method enhanced the performance of the classifier because it can generate rare samples conserving the distribution of the original data for the balancing of the thermal comfort dataset. In addition, the result also shows the weighted loss function for GANs improves the diversity of the generated dataset compare to GANs

without the weighted loss function.

In the latter of this chapter, we proposed a data balancing method for regression datasets, including physiological time-series data. We demonstrated that the estimation accuracy improved for rare cases by balancing datasets, including time-series data measured by physiological sensors. The result indicated that the balancing method helps the effective extraction of physiological time-series features to estimate numerical values. As a limitation, we need a further study on the applicability of our method. TS_SMOTE was proposed as a shape-preserving method to augment time-series. TS_SMOTE works well for acceleration time-series to detect epilepsy and falls. Our work applies TS_SMOTE-based augmentation for time-series of heart rate, body temperature, electrodermal activity, and acceleration to estimate thermal sensation and core temperature. However, the applicability of TS_SMOTE for other tasks has not been shown. When we use TS_SMOTE-based time-series augmentation, it is necessary to observe how the augmented dataset affects the estimator's performance.

Chapter 6

Conclusion

This dissertation has presented a novel approach to correct the fluctuation of the thermal camera's measurement, estimate the thermal sensation using transfer learning, and balance the dataset, including time-series for regression. The primary goal of this dissertation is to realize a thermal sensation estimation method by machine learning approaches. In this dissertation, three primary contributions have been made. The three key steps to build the machine learning model for TSV estimation are the physiological data collection, the selection and construction of the machine learning model, and the training schemes to address imbalanced datasets. The data collection requires accurate sensing to measure physiological data, especially by mobile devices. Another key step is the selection and construction of the machine learning model. The optimization is also critical to build a machine learning model to estimate thermal sensation using physiological data. The dataset collected in uncontrolled environments is usually imbalanced. To deal with the imbalanced dataset is necessary to train the estimator using the dataset collected in such an environment. This dissertation has elaborated on the above three primary steps to achieve our goals as follows.

Firstly, we design and evaluate ThermalWrist, a dynamic offset correction for a smartphone thermal camera using a wristband sensor for low cost and accurate temperature monitoring. The key idea is that the measurement fluctuation of the thermal cameras is caused by the offset error common in all the pixels in a single thermal image. ThermalWrist refers to the temperature to correct the offset error. We selected a palm and a wrist for the reference points and evaluated them by comparing the correction error. Through the real experiment with 876 samples from eight subjects, we confirmed that ThermalWrist remarkably improves three evaluation metrics' accuracy. The limitation of our method depends on the variance of reference point temperature distribution. The experiment in a cool environment shows the variance is larger when the temperature measured by a wristband sensor is lower. The wrist temperature needs to be 32.60 °C or higher in order to utilize our method fully. This result indicates our method is effective in the usual indoor ambient temperature. The result also shows that our method corrected large errors because our method corrected the offset error of a whole thermal image. It is important when we use thermal cameras to capture the change and the abnormal

detection of the body temperature of one-shot observation.

Secondly, we propose thermal sensation estimation using a thermal camera and a wristband sensor. We combine the different estimation models by considering temporal decay of the accuracy to work when a thermal camera can not capture the temperature distribution of the user's face. Experimental data was collected over a year from 15 subjects and was used to make it possible to estimate thermal sensation in all seasons. Through the preliminary study, we confirmed that our method effectively detects the redundant use of air conditioning. After that, we propose TSVNet. This is a deep learning-based approach to combine opportunistic data and time-series data. The key idea is to combine models designed for opportunistic data and time-series data separately by transfer learning. The evaluation result showed that the estimation performance improved in the dynamic environment by utilizing time-series data measured by a wristband sensor. In addition, we also proposed the data balancing method for datasets that contain time-series data. The result indicated that the balancing method effectively extracted features. We also investigated the duration of time before the measurement that affects the current thermal sensation. The result showed that the previous 10 minutes is sufficient for the estimation. As the applicability of our method, the estimator's performance is not enough to control HVAC systems automatically without interactions from occupants. However, the improved estimator will contribute to mitigate stress to occupants for a recommendation of HVAC control, cloth level, and seat positions in offices.

Thirdly, we propose a data balancing method for thermal sensation datasets using conditional Wasserstein GAN with a weighted loss function. We evaluated the performance of our data balancing method using six metrics used in the existing work. The result shows our method enhanced the classifier's performance because it can generate rare samples conserving the distribution of the original data for the balancing of the thermal sensation dataset. In addition, we proposed a data balancing method for regression datasets, including physiological time-series data. We showed that the estimation accuracy improved for rare cases when we used the balanced dataset. The result indicated that our method helps the effective augmentation of physiological time-series features to estimate numerical values. The applicability of our method depends on TS_SMOTE, which was proposed as a shape-preserving method to augment time-series. It was shown that TS_SMOTE works well for acceleration time-series to detect epilepsy and falls in the original work. Our work applies TS_SMOTE-based augmentation for time-series of heart rate, body temperature, electrodermal activity, and acceleration. Because the applicability of TS_SMOTE for other tasks has not been shown, we need to observe how the augmented dataset affects the estimator's performance before using the augmentation.

Through these contributions, an estimator of TSV is constructed based on a machine learning approach. Our study leaves the potentials for further studies for TSV estimation using mobile devices such as smartphone thermal cameras and wearable sensors. For instance, we expect the estimation of TSV to be enhanced by precisely considering human body shape information, such as body mass index and amount of clothes, such as a clo value. Our method can be combined to estimate such values

based on image processing using visible cameras. The improvement of regression performance by data balancing infers the performance of numerical value estimations is affected by a response variable distribution. Therefore, imbalanced regression performance in other domains can also be improved by generating samples appropriately for each situation. Consequently, this dissertation has established the foundation of the thermal sensation estimation method to enhance air control in buildings for all humankind.

Acknowledgement

Through the trials and tribulations of this Ph.D. thesis, I would like to express my sincere appreciation for the continued support of my supervisors, Specially Appointed Professor Teruo Higashino and Professor Hirozumi Yamaguchi of Osaka University. Again I express my heartiest gratitude to them for their encouragement and valuable comments in my learning the posture as a researcher.

I am very grateful to Professor Masayuki Murata, Professor Takashi Watanabe, Professor Toru Hasegawa, and Professor Morito Matsuoka of Osaka University for their valuable comments and suggestions concerning this thesis.

I am heartily grateful to Associate Professor Akira Uchiyama for the precious advice and technical discussions provided throughout our research, anytime day or night.

I would like to thank Associate Professor Akihito Hiromori, Assistant Professor Tatsuya Amano, Specially Appointed Assistant Professor Teruhiro Mizumoto, Erdélyi Viktor, and Rizk Hamada for their frank and valuable comments.

Thanks go to all colleagues of the Mobile Computing laboratory for their feedback, encouragement, and support over six years.

I am also very grateful to the secretaries of the Mobile Computing laboratory for their support.

Finally, I would like to thank my family and friends for their understanding and our idle talking, which helped my gratifying life as a Ph.D. student.

Bibliography

[1] UN. General Assembly (71st sess. : 2016-2017), "Work of the statistical commission pertaining to the 2030 agenda for sustainable development : resolution / adopted by the general assembly," United Nations Digital Library, 2017.

[2] M. Santamouris and K. Vasilakopoulou, "Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation," *e-Prime*, p. 100002, 2021.

[3] O. Seppänen, W. J. Fisk, and Q. Lei-Gomez, "Effect of temperature on task performance in office environment," in *Proceedings of Cold climate HVAC Conference*, July 2006, pp. 74–84.

[4] M. Lee, K.-W. Mui, L. T. Wong, W. Chan, E. W. M. Lee, and C. T. Cheung, "Student learning performance and indoor environmental quality (IEQ) in air-conditioned university teaching rooms," *Building and Environment*, vol. 49, pp. 238–244, March 2012.

[5] Y. Abdelrahman, E. Velloso, T. Dingler, A. Schmidt, and F. Vetere, "Cognitive Heat: Exploring the usage of thermal imaging to unobtrusively estimate cognitive load," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 33:1–33:20, September 2017.

[6] J. Ranjan and J. Scott, "ThermalSense: Determining dynamic thermal comfort preferences using thermographic imaging," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, September 2016, pp. 1212–1222.

[7] M. Burzo, M. Abouelenien, V. Pérez-Rosas, C. Wicaksono, Y. Tao, and R. Mihalcea, "Using infrared thermography and biosensors to detect thermal discomfort in a building's inhabitants," in *Proceedings of the ASME 2014 International Mechanical Engineering Congress and Exposition*. ASME, 2014, pp. 1–11.

[8] H. Genno, K. Ishikawa, O. Kanbara, M. Kikumoto, Y. Fujiwara, R. Suzuki, and M. Osumi, "Using facial skin temperature to objectively evaluate sensations," *International Journal of Industrial Ergonomics*, vol. 19, no. 2, pp. 161–171, 1997, kansei Engineering and Comfort.

[9] A. Basu, A. Routray, S. Shit, and A. K. Deb, "Human emotion recognition from facial thermal image based on fused statistical feature and multi-class svm," in *Proceedings of 2015 Annual IEEE India Conference (INDICON)*, 2015, pp. 1–5.

[10] Empatica, "Real-time physiological signals — E4 EDA/GSR sensor," https://www.empatica.com/en-int/research/e4/, accessed on December 13, 2021.

[11] Empatica, "Embrace2 Seizure Monitoring — Smarter Epilepsy Management — Embrace Watch — Empatica," https://www.4mpatica.com/en-eu/embrace2/, accessed on December 13, 2021.

[12] FLIR, "FLIR ONE Pro — FLIR Systems," https://www.flir.com/flir-one/, accessed on December 13, 2021.

[13] P. W. Kruse, *Uncooled thermal imaging: Arrays, systems, and applications.* SPIE Press (TT51), July 2001.

[14] P. O. Fanger, *Thermal Comfort: Analysis and Applications in Environmental Engineering.* McGraw-Hill, 1970.

[15] D. Teli, M. Jentsch, P. James, and A. Bahaj, "Field study on thermal comfort in a UK primary school," in *Proceedings of 7th Windsor Conference: The Changing Context of Comfort in an Unpredictable World*, April 2012, pp. 1–21.

[16] L. Barrios and W. Kleiminger, "The comfstat - automatically sensing thermal comfort for smart thermostats," in *Proceedings of the 2017 IEEE International Conference on Pervasive Computing and Communications, Hawaii, USA*, March 2017, pp. 257–266.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, June 2002.

[18] L. Torgo, R. Ribeiro, B. Pfahringer, and P. Branco, "SMOTE for regression," *Progress in Artificial Intelligence*, vol. 8154, pp. 378–389, September 2013.

[19] T. Chaudhuri, Y. C. Soh, H. Li, and L. Xie, "Machine learning driven personal comfort prediction by wearable sensing of pulse rate and skin temperature," *Building and Environment*, vol. 170, p. 106615, March 2020.

[20] A. Aryal and B. Becerik-Gerber, "A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor," *Building and Environment*, vol. 160, p. 106223, August 2019.

[21] A. C. Cosma and R. Simha, "Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions," *Building and Environment*, vol. 148, pp. 372–383, January 2019.

[22] E. A. de la Cal, J. R. Villar, P. M. Vergara, Á. Herrero, and J. Sedano, "Design issues in time series dataset balancing algorithms," *Neural Computing and Applications*, vol. 32, pp. 1287–1304, January 2019.

[23] J.-H. Choi and V. Loftness, "Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations," *Building and Environment*, vol. 58, pp. 258–269, December 2012.

[24] B. Tag, G. Chernyshov, and K. Kunze, "Facial temperature sensing on smart eyewear for affective computing," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. New York, NY, USA: ACM, September 2017, pp. 209–212.

[25] N. Kawakami and A. Tsutsumi, "The stress check program: a new national policy for monitoring and screening psychosocial stress in the workplace in japan," *Journal of Occupational Health*, vol. 58, no. 1, pp. 1–6, February 2016.

[26] J. Stemberger, R. S. Allison, and T. Schnell, "Thermal imaging as a way to classify cognitive workload," in *Proceedings of 2010 Canadian Conference on Computer and Robot Vision*, May 2010, pp. 231–238.

[27] C. K. L. Or and V. G. Duffy, "Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement," *Occupational Ergonomics*, vol. 7, pp. 83–94, July 2007.

[28] J. Kang, J. Mcginley, G. McFadyen, and K. Babski-Reeves, "Determining learning level and effective training times using thermography," in *Proceedings of 25th Army Science Conference*, November 2006, pp. 1–6.

[29] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal imaging of physiological stress and its affective potential," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 366–378, November 2012.

[30] C. Puri, L. Olson, I. Pavlidis, J. Levine, and J. Starren, "Stresscam: Non-contact measurement of users' emotional states through thermal imaging," in *Proceedings of the 2005 ACM Conference on Human Factors in Computing Systems*, vol. 2, April 2005, pp. 1725–1728.

[31] S. Ioannou, V. Gallese, and A. Merla, "Thermal infrared imaging in psychophysiology: potentialities and limits," *Psychophysiology*, vol. 51, no. 10, pp. 951–963, June 2014.

[32] I. Pavlidis and J. Levine, "Thermal image analysis for polygraph testing," *IEEE Pulse*, vol. 21, no. 6, pp. 56–64, November 2002.

[33] E. Salazar-López, E. Domínguez, V. J. Ramos, J. de la Fuente, A. Meins, O. Iborra, G. Gálvez, M. A. Rodríguez-Artacho, and E. Gómez-Milán, "The mental and subjective skin: Emotion, empathy, feelings and thermography," *Consciousness and Cognition*, vol. 34, pp. 149–162, July 2015.

[34] L. Gane, S. Power, A. Kushki, and T. Chau, "Thermal imaging of the periorbital regions during the presentation of an auditory startle stimulus," *PLOS ONE*, vol. 6, no. 11, pp. 1–8, November 2011.

[35] S. J. Ebisch, T. Aureli, D. Bafunno, D. Cardone, G. L. Romani, and A. Merla, "Mother and child in synchrony: Thermal facial imprints of autonomic contagion," *Biological Psychology*, vol. 89, no. 1, pp. 123–129, January 2012.

[36] B. Manini, D. Cardone, S. Ebisch, D. Bafunno, T. Aureli, and A. Merla, "Mom feels what her child feels: Thermal signatures of vicarious autonomic response while watching children in a stressful situation," *Frontiers in human neuroscience*, vol. 7, p. 299, June 2013.

[37] A. C. Hahn, R. D. Whitehead, M. Albrecht, C. E. Lefevre, and D. I. Perrett, "Hot or not? thermal reactions to social contact," *Biology letters*, vol. 8, pp. 864–867, May 2012.

[38] P. Nurmi and S. Tarkoma, "Low-cost support for search and rescue operations using off-the-shelf sensor technologies," *Electronics Letters*, vol. 53, pp. 1011–1013, July 2017.

[39] M. L. Mauriello, M. Saha, E. B. Brown, and J. E. Froehlich, "Exploring novice approaches to smartphone-based thermographic energy auditing: A field study," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI '17. ACM, May 2017, pp. 1768–1780.

[40] M. Jaspers, M. Carrière, A. M. de Vries, J. Klaessens, and P. van Zuijlen, "The flir one thermal imager for the assessment of burn wounds: Reliability and validity study," *Burns*, vol. 43, no. 7, pp. 1516–1523, November 2017.

[41] M. Vollmer and K.-P. Möllmann, *Infrared Thermal Imaging: Fundamentals, Research and Applications*. Wiley-VCH, 2011.

[42] T. Malmivirta, J. Hamberg, E. Lagerspetz, X. Li, E. Peltonen, H. Flores, and P. Nurmi, "Hot or not? robust and accurate continuous thermal imaging on flir cameras," in *Proceedings of 2019 IEEE International Conference on Pervasive Computing and Communications*, March 2019, pp. 1–9.

[43] P. Nugent, J. Shaw, and N. J. Pust, "Radiometric calibration of infrared imagers using an internal shutter as an equivalent external blackbody," *Optical Engineering*, vol. 53, p. 123106, December 2014.

[44] D. Mitchell, C. H. Wyndham, and T. Hodgson, "Emissivity and transmittance of excised human skin in its thermal emission wave band." *Journal of Applied Physiology*, vol. 23, no. 3, pp. 390–394, September 1967.

[45] FLIR Systems Japan K.K., "Infrared thermography guidebook for researchers," https://www.flirmedia.com/MMC/THG/Brochures/T559243/T559243_JP.pdf, accessed on December 14.

[46] N. H. Wong and S. S. Khoo, "Thermal comfort in classrooms in the tropics," *Energy and Buildings*, vol. 35, no. 4, pp. 337–351, May 2003.

[47] G. S. Brager and R. J. de Dear, "Thermal adaptation in the built environment: a literature review," *Energy and Buildings*, vol. 27, no. 1, pp. 83–96, February 1998.

[48] S. Takada, H. Kobayashi, and T. Matsushita, "Thermal model of human body fitted with individual characteristics of body temperature regulation," *Building and Environment*, vol. 44, no. 3, pp. 463 – 470, March 2009.

[49] D. Mitchell and C. H. Wyndham, "Comparison of weighting formulas for calculating mean skin temperature," *Journal of Applied Physiology*, vol. 26, no. 5, pp. 616–622, May 1969.

[50] K. Nkurikiyeyezu, K. Shoji, A. Yokokubo, and G. Lopez, "Thermal comfort and stress recognition in office environment," in *Proceedings of the 12th International Conference on Health Informatics*, February 2019, pp. 256–263.

[51] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, *Handbook of Psychophysiology*, 3rd ed. Cambridge University Press, 2007.

[52] N. Gerrett, B. Redortier, T. Voelcker, and G. Havenith, "A comparison of galvanic skin conductance and skin wettedness as indicators of thermal discomfort during moderate and high metabolic rates," *Journal of Thermal Biology*, vol. 38, no. 8, pp. 530–538, August 2013.

[53] Y. P. Sun, N. Zhu, Z. Tian, N. Li, and J. Liu, "Thermal discomfort evaluation in hot environment," *Frontiers of Green Building, Materials and Civil Engineering II*, vol. 193, pp. 1320–1323, August 2012.

[54] A. Ugursal and C. Culp, "Gender differences of thermal comfort perception under transient environmental and metabolic conditions," *ASHRAE Transactions*, vol. 119, pp. 52–62, July 2013.

[55] S. Liu, S. Schiavon, H. P. Das, M. Jin, and C. J. Spanos, "Personal thermal comfort models with wearable sensors," *Building and Environment*, vol. 162, p. 106281, September 2019.

[56] M. H. Hasan, F. Alsaleem, and M. Rafaie, "Sensitivity study for the pmv thermal comfort model and the use of wearable devices biometric data for metabolic rate estimation," *Building and Environment*, vol. 110, pp. 173–183, December 2016.

[57] D. Li, C. C. Menassa, and V. R. Kamat, "Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras," *Applied Energy*, vol. 251, p. 113336, October 2019.

[58] A. C. Cosma and R. Simha, "Using the contrast within a single face heat map to assess personal thermal comfort," *Building and Environment*, vol. 160, p. 106163, August 2019.

[59] W. Hu, Y. Luo, Z. Lu, and Y. Wen, "Heterogeneous transfer learning for thermal comfort modeling," in *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, November 2019, pp. 61–70.

[60] N. Morresi, S. Casaccia, M. Sorcinelli, M. Arnesano, A. Uriarte, J. I. Torrens-Galdiz, and G. M. Revel, "Sensing physiological and environmental quantities to measure human thermal comfort through machine learning techniques," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 12 322–12 337, March 2021.

[61] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–1780, December 1997.

[62] D. Cóstola, G. Carreira, L. O. Fernandes, and L. C. Labaki, "Seasonal thermal sensation vote – an indicator for long-term energy performance of dwellings with no HVAC systems," *Energy and Buildings*, vol. 187, pp. 64–76, March 2019.

[63] M. Giamalaki and D. Kolokotsa, "Understanding the thermal experience of elderly people in their residences: Study on thermal comfort and adaptive behaviors of senior citizens in crete, greece," *Energy and Buildings*, vol. 185, pp. 76–87, February 2019.

[64] M. Jain, A. Singh, and V. Chandan, "Portable+: A ubiquitous and smart way towards comfortable energy savings," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–22, June 2017.

[65] S. Zhu, D. Dalgo, J. Srebric, and S. Kato, "Cooling efficiency of a spot-type personalized air-conditioner," *Building and Environment*, vol. 121, pp. 35–48, August 2017.

[66] Y. Qiao, T. Cao, J. Muehlbauer, Y. Hwang, and R. Radermacher, "Experimental study of a personal cooling system integrated with phase change material," *Applied Thermal Engineering*, vol. 170, p. 115026, April 2020.

[67] F. Barbic, M. Minonzio, B. Cairo, L. Cerina, D. Shiffer, S. Rigo, E. Nappi, A. Bisoglio, P. Verzeletti, F. Badilini, M. Vaglio, M. Santambrogio, R. Gatti, A. Porta, F. Dipaola, and R. Furlan, "Effect of a cool classroom microclimate on symbolic indexes of cardiac autonomic control and cognitive performances in undergraduate students," in *Proceedings of the 11th Conference of the European Study Group on Cardiovascular Oscillations*, July 2020, pp. 1–2.

[68] S. I. Nikolenko, *Synthetic Data for Deep Learning*. Springer, Cham, 2019.

[69] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, April 2018.

[70] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, June 2008, pp. 1322 – 1328.

[71] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," *Advances in Intelligent Computing*, vol. 3644, pp. 878–887, September 2005.

[72] G. Batista, A. Bazzan, and M.-C. Monard, "Balancing training data for automated annotation of keywords: a case study." in *Proceedings of Workshop on Bioinformatics*, January 2003, pp. 10–18.

[73] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, June 2004.

[74] P. Branco, L. Torgo, and R. P. Ribeiro, "SMOGN: a pre-processing approach for imbalanced regression," in *Proceedings of 1st International Workshop on Learning with Imbalanced Domains: Theory and Applications*, September 2017, pp. 36–50.

[75] T. Zhu, C. Luo, J. Li, S. Ren, and Z. Zhang, "Oversampling for imbalanced time series data," https://arxiv.org/abs/2004.06373, April 2020.

[76] H. GM, M. K. Gourisaria, M. Pandey, and S. S. Rautaray, "A comprehensive survey and analysis of generative models in machine learning," *Computer Science Review*, vol. 38, p. 100285, November 2020.

[77] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS'14. Cambridge, MA, USA: MIT Press, December 2014, pp. 2672–2680.

[78] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," https://arxiv.org/abs/1312.6114, December 2013.

[79] Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," https://arxiv.org/abs/2002.12478, February 2020.

[80] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. V. Essen, A. A. S. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, March 2019.

[81] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine Vision and Applications*, vol. 25, no. 1, pp. 245–262, November 2013.

[82] L. E. Yanmaz, Z. Okumus, and E. Dogan, "Instrumentation of thermography and its applications in horses," *Journal of Animal and Veterinary Advances*, vol. 6, no. 7, pp. 858–862, 2007.

[83] P. D. Warriss, S. J. Pope, S. N. Brown, L. J. Wilkins, and T. G. Knowles, "Estimating the body temperature of groups of pigs by thermal imaging," *Veterinary Record*, vol. 158, no. 10, pp. 331–334, March 2006.

[84] C. Vellaichamy, D. S. Jayas, and N. D. White, "Thermal imaging for detecting fungal infection in stored wheat," *Journal of Stored Products Research*, vol. 46, no. 3, pp. 174–179, July 2010.

[85] J. R. M. Dios and A. Ollero, "Automatic detection of windows thermal heat losses in buildings using uavs," in *Proceedings of 2006 World Automation Congress*, June 2006, pp. 1–6.

[86] Y. Aragane and H. Higashino, "The "enseki" sandbath: A novel, safe and effective far-infrared bathing procedure for health," *Photodermatology, Photoimmunology & Photomedicine*, vol. 35, no. 1, pp. 31–39, August 2018.

[87] P. C. Priest, A. R. Duncan, L. C. Jennings, and M. G. Baker, "Thermal image scanning for influenza border screening: Results of an airport screening study," *PLOS ONE*, vol. 6, no. 1, pp. 1–7, January 2011.

[88] I. Pavlidis, N. Eberhardt, and J. Levine, "Human behaviour: Seeing through the face of deception," *Nature*, vol. 415, p. 35, February 2002.

[89] J. F. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 679–698, November 1986.

[90] OpenCV, "Template matching — opencv 2.4.13.7 documentation," https://docs.opencv.org/2.4.13.7/doc/tutorials/imgproc/histograms/template_matching/template_matching.html, accessed on December 14, 2021.

[91] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, December 2001, pp. I–511–I–518.

[92] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, "A fusion approach for efficient human skin detection," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 1, pp. 138–147, February 2012.

[93] A. R. Smith, "Color gamut transform pairs," *SIGGRAPH Comput. Graph.*, vol. 12, no. 3, pp. 12–19, August 1978.

[94] H. C. E. Paul and J. G. Jelle, "Enhancing scatterplots with smoothed densities," *Bioinformatics*, vol. 20, no. 5, pp. 623–628, 2004.

[95] O. Kaynakli and M. Kilic, "Investigation of indoor thermal comfort under transient conditions," *Building and Environment*, vol. 40, no. 2, pp. 165–174, February 2005.

[96] R. Zhao, "Investigation of transient thermal environments," *Building and Environment*, vol. 42, no. 12, pp. 3926–3932, December 2007.

[97] American Society of Heating, Refrigerating and Air-Conditioning Engineers, *Thermal Environmental Conditions for human occupancy.* ASHRAE Standard 55-2017, 2017.

[98] A. Ghahramani, G. Castro, S. A. Karvigh, and B. Becerik-Gerber, "Towards unsupervised learning of thermal comfort using infrared thermography," *Applied Energy*, vol. 211, pp. 41–49, February 2018.

[99] H. Liu, J. Liao, D. Yang, X. Du, P. Hu, Y. Yang, and B. Li, "The response of human thermal perception and skin temperature to step-change transient thermal environments," *Building and Environment*, vol. 73, pp. 232–238, March 2014.

[100] H. Yoshikawa, A. Uchiyama, and T. Higashino, "Dynamic offset correction for smartphone thermal cameras using a wristband sensor," in *Proceedings of the 5th Workshop on Sensing Systems and Applications Using Wrist Worn Smart Devices*, March 2019, pp. 165–170.

[101] H. Zhang, E. Arens, C. Huizenga, and T. Han, "Thermal sensation and comfort models for non-uniform and transient environments: part I: Local sensation of individual body parts," *Building and Environment*, vol. 45, no. 2, pp. 380–388, February 2010.

[102] ——, "Thermal sensation and comfort models for non-uniform and transient environments, part II: Local comfort of individual body parts," *Building and Environment*, vol. 45, no. 2, pp. 389–398, February 2010.

[103] ——, "Thermal sensation and comfort models for non-uniform and transient environments, part III: Whole-body sensation and comfort," *Building and Environment*, vol. 45, no. 2, pp. 399–410, February 2010.

[104] H. Critchley and Y. Nagai, *Electrodermal Activity (EDA)*. Encyclopedia of Behavioral Medicine, Springer, 2013, pp. 666–669.

[105] H. Wang, J. Hu, and W. Deng, "Face feature extraction: A complete review," *IEEE Access*, vol. 6, pp. 6001–6039, March 2018.

[106] H. Yoshikawa, A. Uchiyama, and T. Higashino, "ThermalWrist: Smartphone thermal camera correction using a wristband sensor," *Sensors*, vol. 19, no. 18, p. 3826, September 2019.

[107] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the 3rd international conference on knowledge discovery and data mining*, ser. AAAIWS'94. AAAI Press, July 1994, pp. 359–370.

[108] FLIR, "FLIR T540," https://www.flir.com/products/t540/, accessed on December 14, 2021.

[109] G. Havenith, I. Holmér, and K. Parsons, "Personal factors in thermal comfort assessment: clothing properties and metabolic heat production," *Energy and Buildings*, vol. 34, no. 6, pp. 581–591, July 2002.

[110] T. Ogulata, "The effect of thermal insulation of clothing on human thermal comfort," *Fibres and Textiles in Eastern Europe*, vol. 15, pp. 67–72, April 2007.

[111] P. X. Gao and S. Keshav, "SPOT: A smart personalized office thermal control system," in *Proceedings of the Fourth International Conference on Future Energy Systems*, ser. e-Energy '13. Association for Computing Machinery, January 2013, pp. 237–246.

[112] E. Laftchiev, D. Romeres, and D. Nikovski, "Personalizing individual comfort in the group setting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, May 2021, pp. 15 339–15 346.

[113] L. He, Y. Lee, and K. G. Shin, "Mobile device batteries as thermometers," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 1, pp. 1–21, March 2020.

[114] L. Jiang, X. Lin, X. Liu, C. Bi, and G. Xing, "SafeDrive: Detecting distracted driving behaviors using wrist-worn devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–22, January 2018.

[115] J. Costa, F. Guimbretière, M. F. Jung, and T. Choudhury, "BoostMeUp: Improving cognitive performance in the moment by unobtrusively regulating emotions with a smartwatch," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 2, pp. 1–23, July 2019.

[116] C. L. Castro and A. P. Braga, "Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data," *IEEE transactions on neural networks and learning systems*, vol. 24, no. 6, pp. 888–899, February 2013.

[117] J. Johnson and T. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, p. 27, 03 2019.

[118] M. Mirza and S. Osindero, "Conditional generative adversarial nets," http://arxiv.org/abs/1411.1784, pp. 1–7, November 2014.

[119] M. Quintana, S. Schiavon, K. W. Tham, and C. Miller, "Balancing thermal comfort datasets: We gan, but should we?" in *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, ser. BuildSys '20, November 2020, pp. 120–129.

[120] P. Branco, L. Torgo, and R. P. Ribeiro, "Pre-processing approaches for imbalanced distributions in regression," *Neurocomputing*, vol. 343, pp. 76–99, May 2019, learning in the Presence of Class Imbalance and Concept Drift.

[121] H. Yoshikawa, A. Uchiyama, and T. Higashino, "TSVNet: Combining time-series and opportunistic sensing by transfer learning for dynamic thermal sensation estimation," *IEEE Access*, pp. 102 835–102 846, July 2021.

[122] L. Torgo and R. Ribeiro, "Utility-based regression," in *Knowledge Discovery in Databases: PKDD 2007*. Berlin, Heidelberg: Springer Berlin Heidelberg, September 2007, pp. 597–604.

[123] M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, vol. 27, no. 3, pp. 832–837, September 1956.

[124] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, September 1962.

[125] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, Inc., August 1992.

[126] S. Munir, J. Francis, M. Quintana, N. v. Frankenberg, and M. Bergés, "Dataset: Inferring thermal comfort using body shape information utilizing depth sensors," in *Proceedings of the 2nd Workshop on Data Acquisition To Analysis*, ser. DATA'19, November 2019, pp. 13–15.

[127] P. Jayathissa, M. Quintana, M. Abdelrahman, and C. Miller, "Humans-as-a-sensor for buildings—intensive longitudinal indoor comfort models," *Buildings*, vol. 10, no. 10, p. 174, August 2020.

[128] V. F. Ličina, T. Cheung, H. Zhang, R. de Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G. Brager, P. Li, and S. Kaam, "ASHRAE Global Thermal Comfort Database II," Dryad, Dataset, July 2018.

[129] cosinuss°, "cosinuss° one – performance monitoring," https://www.cosinuss.com/en/products/one/, accessed on December 13, 2021.

[130] UNION TOOL, "Whs-3 wearable heart rate sensor," https://www.uniontool-mybeat.com/SHOP/8600085.html, Accessed on April 29, 2021 (in Japanese).

[131] L. Torgo and R. Ribeiro, "Precision and recall for regression," *Discovery Science*, vol. 4702, pp. 332–346, October 2007.