

Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras

Da Li, Carol C. Menassa*, Vineet R. Kamat

Department of Civil and Environmental Engineering, University of Michigan, United States

HIGHLIGHTS

- Characteristics of the non-intrusive thermal comfort detection are defined.
- A camera network is introduced to assess thermal comfort in multi-occupancy spaces.
- Thermal and RGB-D cameras are fused to measure facial skin temperature.
- Subjects can have flexible postures and movements during the data collection.
- Facial mean skin temperature can serve as an indicator of one's thermal comfort.

ARTICLE INFO

Keywords:

Thermal comfort assessment
Non-intrusive data collection
Infrared thermography
Camera network
Physiological measurements

ABSTRACT

About 40% of the energy produced globally is consumed within buildings, primarily for providing occupants with comfortable work and living spaces. However, despite the significant impacts of such energy consumption on the environment, the lack of thermal comfort among occupants is a common problem that can lead to health complications and reduced productivity. To address this problem, it is particularly important to understand occupants' thermal comfort in real-time to dynamically control the environment. This study investigates an infrared thermal camera network to extract skin temperature features and predict occupants' thermal preferences at flexible distances and angles. This study distinguishes from existing methods in two ways: (1) the proposed method is a non-intrusive data collection approach which does not require human participation or personal devices; (2) it uses low-cost thermal cameras and RGB-D sensors which can be rapidly reconfigured to adapt to various settings and has little or no hardware infrastructure dependency. The proposed camera network is verified using the facial skin temperature collected from 16 subjects in a multi-occupancy experiment. The results show that all 16 subjects observed a statistically higher skin temperature as the room temperature increases. The variations in skin temperature also correspond to the distinct comfort states reported by the subjects. The post-experiment evaluation suggests that the networked thermal cameras have a minimal interruption of building occupants. The proposed approach demonstrates the potential to transition the human physiological data collection from an intrusive and wearable device-based approach to a truly non-intrusive and scalable approach.

1. Introduction

Buildings consume approximately 40% of the energy produced globally resulting in significant impacts on critical non-renewable resources and climate change [1,2]. At the same time, heating, ventilation, and air conditioning (HVAC) systems represent the biggest energy end use in buildings, which account for 48% of the total energy required to operate residential and commercial buildings [3]. However, despite the significant impact of such energy consumption on the

environment, the lack of thermal comfort among building occupants is a common problem where studies reveal that up to 43% of occupants are dissatisfied with the thermal environment in their workplace [4].

The importance of thermal comfort cannot be overemphasized. Several studies have suggested that satisfying thermal environments can lead to a reduced number of complaints, absenteeism, and improved work productivity [5]. On the other hand, it is also not surprising that thermal comfort is an influential factor on occupants' health and well-being, especially given that people spend more than

* Corresponding author at: Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, United States.

E-mail addresses: dli@umich.edu (D. Li), menassa@umich.edu (C.C. Menassa), vkamat@umich.edu (V.R. Kamat).

90% of their time indoors [6]. For example, the reports of sick building syndrome symptoms, such as headaches, eye and throat irritation, have been found to be correlated with the high room temperature [7]. Therefore, optimizing the HVAC operation to improve human satisfaction, health, and energy efficiency can lead to significant social, economic, and environmental benefits.

To this end, a good understanding of occupants' thermal comfort is much needed as it provides useful insights to control the HVAC systems. However, assessing occupants' thermal comfort is not an easy task. Thermal comfort is defined as "the condition of mind which expresses satisfaction with the thermal environment" [8], which implies that it is determined by one's subjective assessment of the environmental condition (e.g., air temperature, relative humidity). In addition, an individual's thermal comfort is also significantly affected by personal conditions consisting of physiological (e.g., gender, body mass index), psychological (e.g., expectation, stress), and behavioral factors (e.g., clothing and activity level) [9–12]. As a result, both personal and temporal variations should be considered to achieve a robust thermal comfort assessment.

To date, researchers have explored different methods to assess thermal comfort. The most well-known approach is the Predicted Mean Vote (PMV) model, which is based on the thermal transfer of the human body and the environment [13]. The PMV model considers four environmental factors (air temperature, mean radiant temperature, air speed, relative humidity) and two human factors (clothing level and metabolic rate) affecting one's thermal sensation and comfort. Later, the adaptive comfort models were proposed to account for adaptive behaviors from occupants (e.g., open the window) to maintain thermally comfortable states in naturally ventilated environments and suggested a wider comfort zone than that of the PMV model [14]. Recently, more attention has been paid to personal comfort models [9,15–18]. In this case, "personal" indicates that the predictive model was entirely and exclusively developed based on one's subjective thermal comfort states under various environmental and human conditions. As a result, the training data of personal comfort models typically vary from one person to another, which reflects occupants' uniqueness and subjectivity in evaluating thermal comfort. Regarding the model training, different machine learning methods can be applied and tuned to choose the best model that yields the highest accuracy in the cross-validation. For example, Li et al. [9] collected environmental and human data (e.g., skin temperature, heart rate) from seven office employees using environment sensors and wristband bio-sensors, and trained personal comfort model for each occupant using the Random Forest classifier, which showed an 80% accuracy in predicting a three-point thermal preference. In general, personal comfort models demonstrate a better predictive power than the PMV and adaptive comfort models as they are highly customized to account for the characteristics of each occupant [17]. In addition, understanding personal thermal comfort provides insights into designing the HVAC control algorithms to optimize overall comfort in a multi-occupancy setting [9] and also enables personalized comfort zones for each individual (e.g., through local HVAC units) in a large space when coupling with occupant positioning systems [16]. From the energy's perspective, personal comfort profiles can be incorporated into the building energy models or agent-based models for detailed energy simulation and occupant feedback (e.g., [19–22]).

To apply personal comfort models, the environmental condition within the proximity of an occupant (also known as the micro-climate around a person) and/or actionable human physiological data, such as skin temperature, heart rate, and respiration rate, are often required as input variables of the model. However, these input variables are generally measured in an "intrusive" way in existing studies. The intrusiveness of data collection comes from (1) the continuous requirement of human participation for real-time thermal comfort evaluation (e.g., [23]); (2) the dependence on wearable devices or personal equipment (e.g., [24,25]); and (3) the requirement of occupants to remain at a

static posture or refrain from excessive body movements (e.g., [26]). The intrusiveness resulting from conventional methods is a significant limitation as it is distracting during regular work time and also impractical in operational built environments.

Therefore, to leverage personal comfort models and address the intrusiveness of conventional data collection methods, this study proposes a non-intrusive and low-cost thermal and RGB-D cameras network to simultaneously assess multiple occupants' thermal comfort states in real multi-occupancy environments with minimal interruption of building occupants. Also, this networked system has high scalability and flexibility and can be easily applied to various built environments.

The paper is organized to first provide a comprehensive review of existing studies on the related topics, which is followed by a discussion of the anticipated characteristics of a non-intrusive data collection approach in Section 2. Then, the technical details of the proposed approach are explained in Section 3. The protocol of data collection experiments is described in Section 4. The experimental results and capabilities of the proposed approach are discussed in Section 5. Finally, Section 6 concludes this study.

2. Background

2.1. Related work

Researchers have explored various approaches and tools to measure the input variables of personal comfort models. In general, existing data collection approaches can be categorized into two main categories: (1) environment oriented comfort assessment, and (2) human physiological measurement oriented comfort assessment.

The environment oriented comfort assessment aims to collect various environmental factors within the proximity of an occupant and then map these factors to his/her self-reported thermal comfort conditions. For example, Feldmeier and Paradiso [23] used wearable sensor nodes to monitor the local temperature and humidity surrounding an occupant and developed decision boundaries based on these two parameters using the Fisher discriminant analysis. Jazizadeh et al. [27] developed a phone application through which occupants can provide their thermal preferences. Personal fuzzy predicted models were developed to correlate each occupant's preference and the ambient air temperature. Recently, Kim et al. [17] designed a comfort chair which can deliver personal heating and cooling through the heat strips and fans mounted on the chair. The control behaviors (i.e., requests to heat and cool) received by the chair and several environmental conditions measured by sensors were included in the personal comfort models for thermal preference prediction. However, these approaches generally fail to consider human physiological data which have been proved as useful indicators of one's thermal comfort state. Lacking such human data can lead to sub-optimal predictions due to the varying human physiological, psychological, and behavioral conditions. For example, occupants can have distinct comfort levels with an extra layer of clothing under the same environmental conditions, which implies the need to leverage human data for robust comfort assessment.

On the other hand, researchers also explored human data in personal comfort models to address the aforementioned limitation, which leads to the physiological measurement based comfort assessment. Among various human physiological data, skin temperature has drawn significant attention due to its strong correlation with thermal comfort. According to thermoregulation mechanisms, the human body maintains its core internal temperature at around 37 °C. When thermoreceptors detect heat or cold stress, the hypothalamus will control body muscles, organs, and nervous system to maintain the homeostasis state [11]. In this process, the human body adjusts heat production and also controls heat loss through the vasodilation and vasoconstriction, which will cause variations in skin temperature. Therefore, skin temperature is commonly adopted as a proxy of the thermoregulation.

To leverage the skin temperature as an effective indicator of thermal

comfort, prior studies have proposed various approaches to continuously measure one's skin temperature, which include (1) wearable sensors or thermocouples that directly contact the skin surface; and (2) infrared thermometers (spot pyrometer) or thermal imaging cameras that can infer non-contact temperature from the infrared radiation. For example, Li et al. [9,18] adopted a wrist-worn fitness tracker to continuously measure occupants' wrist skin temperature and other physiological variables under different thermal conditions. This study compared baseline models which are trained only using the environmental data with those that use both environmental and human physiological data. Results suggested that incorporating human data can significantly improve the predictive power of personal comfort models.

To reduce the intrusiveness resulting from the wearable sensors, infrared thermometers and cameras are ideal tools as they can infer the temperature in a non-contact way. For example, Ghahramani et al. [24] installed infrared thermometers on an eyeglass frame to collect the skin temperature of the front face, cheek, nose, and ear regions and observed statistically significant differences in skin temperature under heat and cold stress conditions.

Compared to infrared thermometers, thermal camera (or infrared imaging cameras) can capture a full thermal image consisting of temperature values of each pixel (as opposed to a single spot measurement from infrared thermometers) and measure objects' temperature from a longer distance [28]. The measurements are also invariant to ambient illumination. In existing studies, researchers used thermal cameras in different ways to model occupants' thermal preference. For example, Ranjan and Scott [29] used the FLIR A655sc camera to take thermographic images of occupants twice per day during a five-week data collection period. Each thermal image is manually labeled offline to identify the skin temperature of different body regions (e.g., palm, forehead), which were then correlated with the reported overall thermal preference. To continuously collect an occupant's skin temperature for real-time analysis, Metzmacher et al. [30] proposed a face and body tracking method to extract the skin temperature of different facial regions in each thermographic image collected from a FLIR A35 camera. However, commodity thermal cameras are generally expensive (over \$5000 as mentioned in the two examples) and cannot be directly incorporated in the building management system due to their large sizes and compatibility issues, which significantly limits their applications in built environments.

To overcome this limitation, previous work by the authors [31] explored a low-cost thermal camera (FLIR Lepton, cost: \$200, dimension: $8.5 \times 11.7 \times 5.6$ mm) to assess thermal comfort through the skin temperature of six facial regions (e.g., forehead, nose). The proposed method can automatically and continuously detect human faces, measure the skin temperature of each facial region, clean and process raw data, and interpret thermal comfort using personal comfort models. Results of this study suggested an 85% accuracy in predicting a three-point thermal preference. More importantly, this pilot study verified the feasibility of incorporating low-cost thermal cameras into the HVAC or building management systems to understand thermal comfort in real time.

Besides infrared thermography, researchers also explored non-contact approaches to measure other physiological factors, such as heart rate, blood perfusion, and respiration activity. For example, Kwon et al. [32] applied the independent component analysis and fast Fourier transform on images captured by regular RGB cameras to detect subtle changes in skin color, which indicates the frequency of cardiac pulses. Similarly, Jung and Jazizadeh [33] compared the facial skin color under cold and hot environments using the Eulerian Video Magnification algorithm. Significant color differences were observed, which suggests variations in the blood perfusion. In another study, Jung and Jazizadeh [34] measured movements of the chest and abdomen using the Doppler radar as a proxy of respiration activities and observed a higher respiration intensity in high temperature environments. However, these studies require the subjects to remain at a static posture as

body movements can affect colors and distances measured by the RGB camera and the Doppler radar.

2.2. Characteristics of non-intrusive thermal comfort assessment

Although the related studies significantly contribute to the current body of knowledge in assessing thermal comfort, these approaches are generally limited to experimental conditions due to the varying degrees of intrusiveness on occupants. Considering the limitations of each reviewed study, a truly non-intrusive approach which can be readily adopted in real operational multi-occupancy environments should have specific characteristics summarized as follows:

- The approach should continuously collect human physiological data for real-time and robust thermal comfort assessment. Therefore, approaches that require manual data collection and offline processing fail this criterion.
- The collection of human physiological data should not require wearable devices, personal equipment, or excessive human feedback, which can cause discomfort, inconvenience or interruptions of building occupants.
- Occupants can have flexible and relaxed postures and possibly move around in the built environment. In other words, occupants are not required to remain at a static posture while the approach is in operation.
- The approach should have high scalability potential such that it can be flexibly configured to various built environments, especially in multi-occupancy spaces where multiple occupants' thermal comfort can be simultaneously assessed without incurring additional adjustments. Typically, studies that require personal equipment fail this criterion due to their hardware dependency. For example, each occupant has to wear a wristband or use the phone app for data collection, which can be cumbersome in large multi-occupancy spaces. If some occupants do not have access to such devices, their comfort preferences are not taken into account.
- The approach should be robust against variations in ambient conditions, such as the lighting intensity. Studies that rely on analyzing different channels of RGB images may fail this criterion due to their sensitivity to lighting variations. For example, a dimmer room or background reflection can substantially change the value of each pixel in an image.

In summary, Table 1 shows the list of significant characteristics of a non-intrusive approach for thermal comfort assessment and Table 2 summarizes the limitations of each related study reviewed in Section 2.1.

Therefore, the objective of this study is to develop a truly non-intrusive physiological sensing approach which satisfies the five main requirements as summarized in Table 1. To this end, this study proposes a camera-occupant network consisting of multiple low-cost thermal and RGB-D cameras to achieve good viewing coverage of the environment and simultaneous interpretations of thermal comfort in operational multi-occupancy built environments. The main challenges of the proposed approach, such as the network formation, camera registration, and skin temperature extraction, are detailed in Section 3.

Table 1
Characteristics of the problem statement.

Serial number	Characteristics of the problem statement
1	Continuous collection of human physiological data
2	Does not require personal devices or human feedback
3	Flexible postures or body movements
4	High scalability in multi-occupancy scenarios
5	Robust against variations in ambient conditions

Table 2
Limitations of existing studies in thermal comfort assessment.

Sources	Devices	Input variables	Unaddressed characteristics
Feldmeier and Paradiso [23]	Wearable sensor nodes (wrist, neck)	Local temperature and humidity	1, 2, 4
Jazizadeh et al. [27]	Phone application	Room temperature and humidity	1, 2, 4
Kim et al. [17]	Personal comfort chair	Control behavior, indoor and outdoor environment factors	1, 2, 4
Li et al. [9,18]	Wrist-worn fitness tracker, phone application	Skin temperature, heart rate, activity level, clothing level, indoor and outdoor environmental factors	2, 4
Ghahramani et al. [24]	Infrared thermometers on eyeglass frame	Skin temperature of the front face, cheek, nose, and ears	2, 4
Ranjan and Scott [29]	FLIR A655sc thermal camera	Skin temperature of multiple body regions	1, 3, 4
Metzmacher et al. [30]	FLIR A35 thermal camera	Skin temperature of multiple facial regions	3, 4
Li et al. [31]	FLIR Lepton 2.5 (low-cost)	Skin temperature of forehead, nose, cheeks, mouth, ears, and neck.	3, 4
Kwon et al. [32]	RGB camera	RGB videos to infer heart rate	3, 4, 5
Jung and Jazizadeh [33]	RGB camera	RGB videos to infer blood perfusion	3, 4, 5
Jung and Jazizadeh [34]	Doppler radar	Chest and abdomen movement to infer respiration rate	3, 4

3. Technical approach

The networked camera system leverages a range of cutting-edge techniques, such as (1) deep neural network to robustly detect multiple occupants from different angles and distances; (2) computer vision to register and track occupants in different camera views and stitch indoor scenes; (3) signal and data processing approaches to remove outliers and smooth the data; and (4) machine learning to develop personal comfort models using skin temperature data. Fig. 1 presents an overview of the thermal camera network.

The thermal camera network is proposed based on the authors' prior work in Li et al. [31] which verified the feasibility of a single low-cost thermal camera for thermal comfort assessment. However, several limitations from Li et al. [31] should be acknowledged: First, the single thermal camera is placed in front of the occupant, which is not suitable for multi-occupancy spaces due to the camera's limited field of view. Second, the thermal camera alone may fail to detect profile faces or frontal faces at a long distance due to its low image resolution. As a result, high resolution RGB cameras are incorporated to achieve robust face detection. Third, the thermal camera is placed at a fixed distance (1 m) from the occupant. As an influential factor of the thermal camera's measurement, occupants' distances to the camera can vary over time in real operational settings. Thus, the viewing distance should be accounted for when measuring skin temperature. To address these

limitations, several improvements are made in this current study (summarized in Table 3). Details in Table 3 are explained in the following subsections which are organized to first introduce the component of the camera network – that is, a single camera node; followed by descriptions of the camera-occupant network including its graph abstraction, occupant registration, data communication, data cleaning, and feature extraction.

3.1. Thermal and RGB-D dual camera system

The camera network shown in Fig. 1 consists of multiple camera nodes that are placed at different locations in a built environment. Each camera node is a low-cost dual camera system comprised of a FLIR Lepton 2.5 thermal camera module and a Microsoft Kinect (an RGB-D camera). As shown in Fig. 2, the thermal camera is rigidly mounted on top of the Kinect.

This dual camera system performs a synergistic function where the Kinect will implement the human detection (by its RGB camera) and provide distance information (by its depth sensor) to register the dual camera system and complement temperature measurements taken at different distances by the thermal camera. Table 4 shows the specifications of these two cameras. Previous work by the authors conducted a preliminary study which demonstrated that the low-cost thermal camera can achieve an acceptable accuracy to interpret thermal

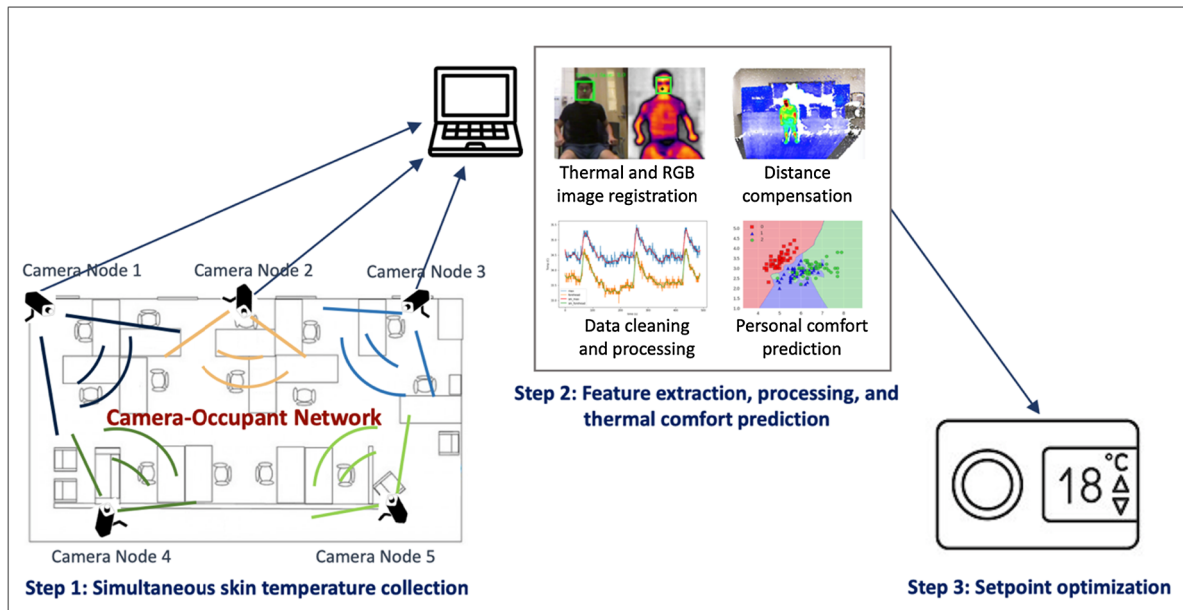


Fig. 1. Overview of the proposed thermal camera network.

Table 3

Comparison of the proposed thermal camera network and the authors' prior work Li et al. [31].

Descriptions	Prior work by the authors [31]	This study
Intended application	Single occupancy	Multi-occupancy
Type of camera	Thermal camera only	Fusion of thermal and RGB-D cameras
Number of cameras	Single camera node	Multiple camera nodes
Face detection approach	Haar cascade algorithm	Can apply various state-of-the-art algorithms
Camera placement	In front of the occupant	Flexible camera placement
Occupant posture and body movement	Refrain from large movements (avoid being out of the camera view)	No constraints
Camera-occupant Distance	Not considered	Compensated for distances
Feature selection	Local skin temperature features from each facial region	Global skin temperature features from the whole facial area

**Fig. 2.** The dual camera system (thermal camera coupled with Microsoft Kinect).**Table 4**

Specifications of the thermal camera and Microsoft Kinect.

FLIR Lepton 2.5	Dimensions	8.5 × 11.7 × 5.6 mm
	Resolution	80 (h) × 60 (v) pixels
	Thermal sensitivity	< 50 mK
	Accuracy	± 5 °C or ± 5% of reading in the working range
	Field of view	51° (h) and 42° (v)
	Price	\$199
Microsoft Kinect	RGB camera resolution	640 (h) × 480 (v) pixels
	Field of view	57° (h) and 43° (v)
	Effective range of depth sensor	0.8–5 m
	Depth accuracy	± 4 cm at the maximum working range (5 m)
	Price	\$48

comfort. For details, please refer to Li et al. [31].

3.1.1. Kinect face detection

The human face is selected as the region of interest as it has a higher density of blood vessels where the variations in skin temperature are more significant due to vasodilation and vasoconstriction [35]. In addition, the human face is generally not covered by clothing such that the emitted infrared energy can be directly received by the thermal camera. In this study, thermal reflection on the human face is not considered under the assumption that the diffuse reflection by the skin surface will not significantly affect the measurement. To detect human faces in thermal images, prior work by the authors tested the Haar Cascade algorithm [36], Histogram of Oriented Gradients [37], and Eigenfaces [38] and found that the Haar Cascade algorithm to be the only feasible method to detect frontal faces as only certain edges (e.g., nose) are preserved in the low-resolution images (80 by 60) [31]. In terms of profile faces, however, none of these methods can robustly detect faces due to the blurred edges, which limits the application of such single thermal camera system in the real built environment as occupants' poses and locations can be very flexible over time. To address this limitation, in this study the authors adopted a Kinect (RGB-D camera) to assist the thermal camera in the face detection. Kinect is

suitable for this task as the high-resolution RGB images (640 by 480) contain more color information than thermal images and thus support more advanced face detection algorithms. In this study, the authors adopted the deep neural network (DNN) based face detectors implemented in the OpenCV library (OpenCV 3.3). Based on the authors' preliminary test, the DNN detector demonstrates a much higher accuracy than the Haar cascade algorithm in both frontal and profile face detections at distances between 0.8 and 5 m, distances typically encountered in indoor environments. However, other algorithms such as the Faster R-CNN and DeepFace (e.g., [39]) can also be applied to assist the thermal camera to locate human faces.

3.1.2. Occupant tracking in a single dual camera node

As more than one occupant can be observed by a single camera node in a multi-occupancy environment, the authors implemented the centroid tracking algorithm to track occupants across image frames. This algorithm assumes that centroids of the same object in the two consecutive frames will have the closest distance [40]. In the first frame, the centroid of each face can be detected using the face detection algorithm introduced in Section 3.1.1. In the subsequent frame at time $t + 1$, the Euclidean distances between each pair of centroids in the current frame $t + 1$ and the previous frame t are calculated and each occupant in frame $t + 1$ is assigned to the ID of its closed centroid in the previous frame t (see Eq. (1)).

$$\text{Occupant ID} = \underset{m \in M_t}{\operatorname{argmin}} \|x_{t+1} - m\| \quad (1)$$

where M_t is a set of the centroids of all subjects at time t ; m is the centroid of one subject in the set M_t ; x_{t+1} is the centroid of a subject at time $t + 1$ (which needs to be updated), and $\|\cdot\|$ is the L_2 -norm. This recursive subject tracking process is illustrated in Fig. 3 with an example of two occupants. As shown in Fig. 3a, at time $t + 1$ the unknown centroids i and j (denoted in triangles) are updated based on their closest Euclidean distances to the centroids at time t (denoted in circles). The unknown centroids at time $t + 2$ (denoted in squares) are updated accordingly (Fig. 3b).

3.1.3. Kinect and thermal camera registration

As the Kinect and thermal camera have different fields of view and resolutions, these two cameras should be registered to find point correspondences such that face coordinates detected in the RGB images can be mapped to thermal images. In this case, both Kinect and thermal camera can be modeled as a pinhole camera which projects the 3D world scene into a 2D image plane through the perspective transformation shown in Eq. (2) [41,42].

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2)$$

Alternatively, in a more concise form,

$$sm = K[R|T]M$$

where M is a 4×1 vector representing the homogeneous coordinate of

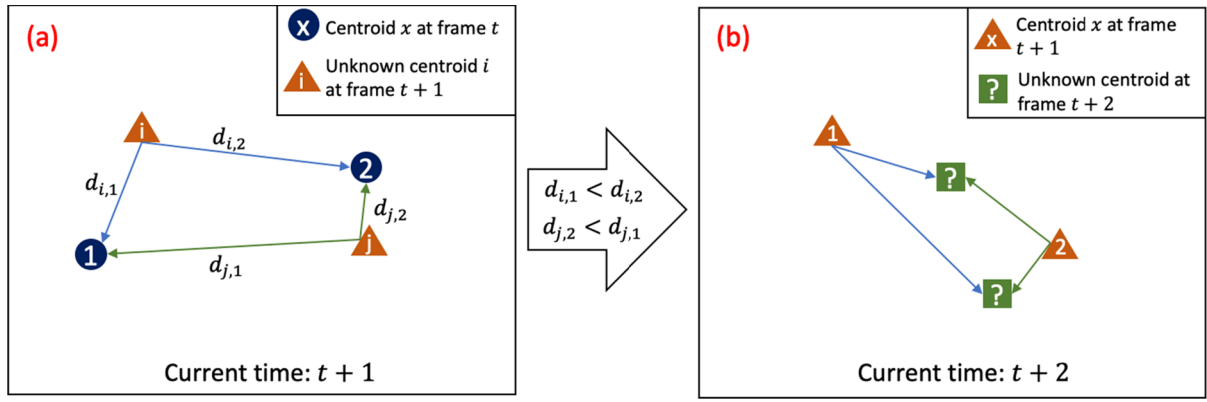


Fig. 3. Centroid tracking algorithm: (a) update at time $t + 1$; (b) update at time $t + 2$.

a 3D point in the world coordinate space; m is a 3×1 vector representing the homogeneous coordinate of a 2D point in the image coordinate; K is the 3×3 intrinsic matrix of the camera consisting of the focal lengths (f_x, f_y) and principal points (C_x, C_y); $[R|T]$ is the 3×4 extrinsic matrix consisting of a rotation R and a translation T ; and s is a scaling factor.

In the dual camera system, the registration process is to estimate the intrinsic matrix of the thermal camera K_{IR} , the intrinsic matrix of the RGB camera K_{RGB} , and the homogeneous transformation matrix $[R|T]$ between the thermal and RGB cameras (see Fig. 4). Once these three unknown matrices are estimated, the point correspondences in two cameras (e.g., (u_1, v_1) and (u_2, v_2)) can be determined according to the pinhole camera model in Eq. (2). In practice, such a dual camera system can be calibrated using the stereo vision calibration process. As shown in Fig. 4, the calibration requires both cameras to observe a planar and predefined pattern, such as a checkerboard or a square grid, from at least two different orientations to determine the unknowns using the maximum likelihood estimation [43].

However, thermal cameras typically cannot detect the regular black-and-white calibration patterns printed on the paper as the infrared energy emitted is uniform across the patterns. Therefore, the authors made a special 6×7 checkerboard pattern from the aluminum foil (in silver) and the vinyl polymer (in black) (Fig. 5a). Each black or silver square has a dimension of 62.5 mm. Due to the color differences, the checkerboard pattern can be detected by the RGB camera to extract corner points (Fig. 5a and c). On the other hand, as the aluminum foil has a higher emissivity, it emits more infrared energy and thus looks

brighter in thermal images. As shown in Fig. 5b and d, the checkerboard corner points can be easily observed by a thermal camera, especially when the pattern is heated up by a hair dryer.

The authors captured twenty pairs of RGB and thermal images from different orientations relative to the dual camera system and implemented the calibration using the Matlab Stereo Camera Calibrator [44]. Although the thermal camera has a low resolution, results showed that the re-projected points are close to the detected points (Fig. 6a), and the mean re-projection error is 1.02 pixels (Fig. 6b), which is acceptable for this application. It is also worth noting that the three unknown matrices only depend on the intrinsic properties of two cameras and their relative pose (how thermal camera is mounted), which are not affected by the distance between the camera and the checkerboard pattern. As a result, the registration process only needs to be done once when configuring the dual camera system.

Fig. 7 shows this result of dual camera registration where the face in the thermal image (labeled in the bounding box, see Fig. 7c) is located based on the face coordinates detected in the RGB image (Fig. 7b) and its corresponding depth data from the Kinect (Fig. 7a).

3.1.4. Distance calibration of the thermal camera

The infrared energy reaching the thermal camera is affected by the distance between the camera and the object surface [45]. Thus, the real-time distance of the camera and each occupant should be measured to fuse the skin temperature data collected from multiple camera nodes. To calibrate distances, the authors conducted a pilot experiment using the low-cost thermal camera to collect mean temperature of the frontal face at different distances from 0.8 m to 2 m with a step size of 0.05 m (room temperature: 26.0 °C, relative humidity: 28.5%). The facial mean temperature was calculated by averaging the measurements of all pixels that exceed a predefined threshold (e.g., measurements below 27 °C were excluded) within the bounding box. The distance was calculated from the point cloud produced by the Kinect. The depth measurement of the Kinect has an accuracy of ± 4 cm within its working range of 5 m which is considered sufficient for this study [46]. At each distance, three thermal images of a subject's frontal face were collected and averaged to represent the measurement at that distance. The whole experiment is conducted within one minute thus the facial skin temperature can be assumed constant during this short period. This calibration experiment was repeated on five different subjects to quantify the impact of distances on temperature measurements and the averaged slope was retained. It is worth noting that the viewing angle is fixed during the calibration (i.e., perpendicular to the face). However, adjusting the angles also affects the facial regions observed by the camera, which will lead to different temperature measurements. For example, an image capturing the frontal face can have a higher measurement than that of a profile face as the forehead region (which has a high temperature) is well captured in the former image.

As shown in Fig. 8a, a linear relationship can be observed from the

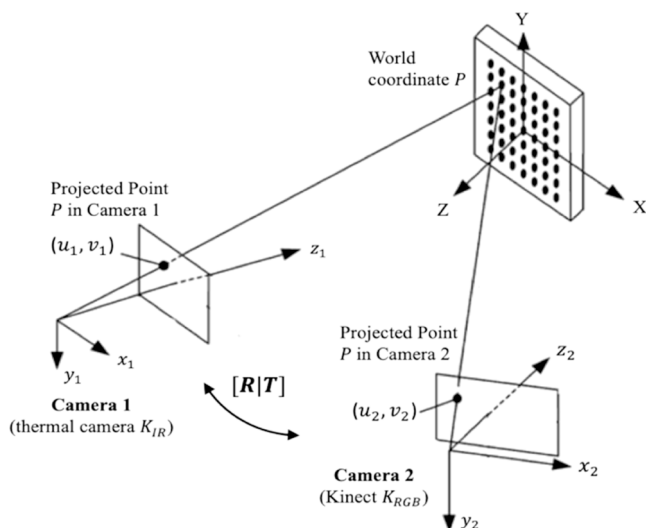


Fig. 4. Stereo vision calibration for Kinect and thermal camera registration.

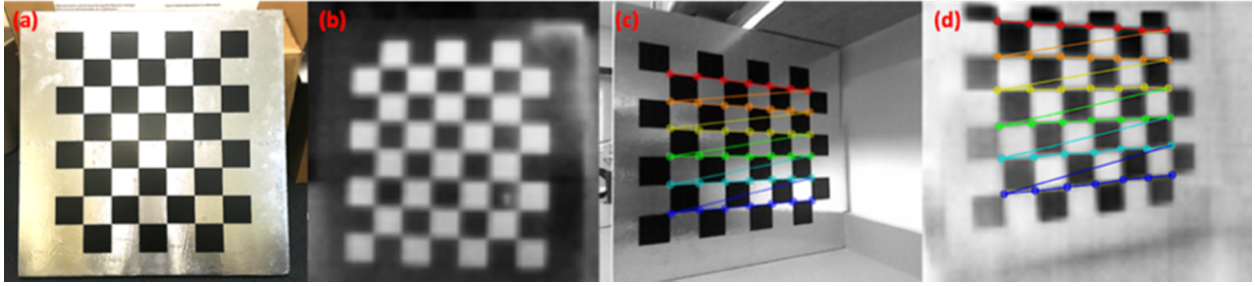


Fig. 5. (a) Special checkerboard for the dual camera registration; (b) Thermal image of the checkerboard (bright squares are regions with higher infrared energy); (c) Corner detection using the Kinect; (d) Corner detection using the thermal camera.

samples ($\hat{y} = -0.50x + 35.02$, adjusted R-square = 0.96, RMSE: 0.04), which implies that skin temperature measurements will drop by 0.5°C for every one-meter increase in distance for the low-cost thermal camera. Fig. 8b shows the residual plots of the linear fit. It can be seen that the residuals are symmetrically distributed around zero and no clear patterns are observed, which also indicates a good fit of the linear model.

3.2. The camera-occupant network

The camera-occupant network is proposed to non-intrusively and simultaneously interpret occupants' thermal comfort in real multi-occupancy spaces. Specifically, this network aims to achieve comprehensive coverage of the environment such that all the occupants can not only be seen by cameras but also have flexible postures and movements during the data collection. It should also be noted that cameras in the network are not placed in front of faces like Li et al. [31] and the images are discarded immediately after retrieving skin temperature data to address the privacy concerns that may arise from any camera systems. This sub-section is organized to first introduce the configuration of the camera network, including its graph abstraction, occupant registration of different camera views, and the data communication between camera nodes.

3.2.1. Graph abstraction of the network

The camera-occupant network is an observation system which contains multiple camera nodes to observe one or more occupants from arbitrary angles and distances. This network can be represented in a graph abstraction adapted from Feng et al. [47] consisting of *nodes* and *edges*. Fig. 9 is an example network which contains three subjects and

two camera nodes. In this graph abstraction, there exist three types of nodes including a *camera node* (denoted as two triangles bounded by a rectangle which represent the dual camera system discussed in Section 3.1), an *occupant node* (denoted as a square), and a *world coordinate node* which represents the origin in the 3D world (denoted as a circle). In addition, there are two types of edges connecting each pair of nodes including *observations* (denoted as solid lines) and *constraints* (denoted as dashed lines). The observation represents the pose between a camera node and an occupant node which can vary over time as occupants change their posture or move around. The observation edges can be estimated using the pinhole camera model introduced in Section 3.1.3. As shown in Fig. 9, occupant nodes 1 and 2 are both observed by camera nodes 1 and 2, while occupant node 3 is only observed by camera node 2.

On the other hand, the constraint edges represent a known geometric relationship within a dual camera system (i.e., the relative pose between the thermal camera and the Kinect) or between dual camera systems (i.e., the relative pose between two dual camera systems mounted in an environment). The constraints can be determined when the camera network is configured up front through the calibration process introduced in Section 3.1.3 and will not be affected by the number, poses, and locations of occupants in the environment. In addition, the origin in the 3D world can be assigned to the location of a camera node and thus their relationship is also represented as a constraint.

To scale up the camera-occupant network in a large multi-occupancy space, more camera nodes can be added to the network. The camera network will be configured such that any occupant node is connected to at least one camera node. Therefore, the skin temperature of each occupant is guaranteed to be collected. To achieve this, the

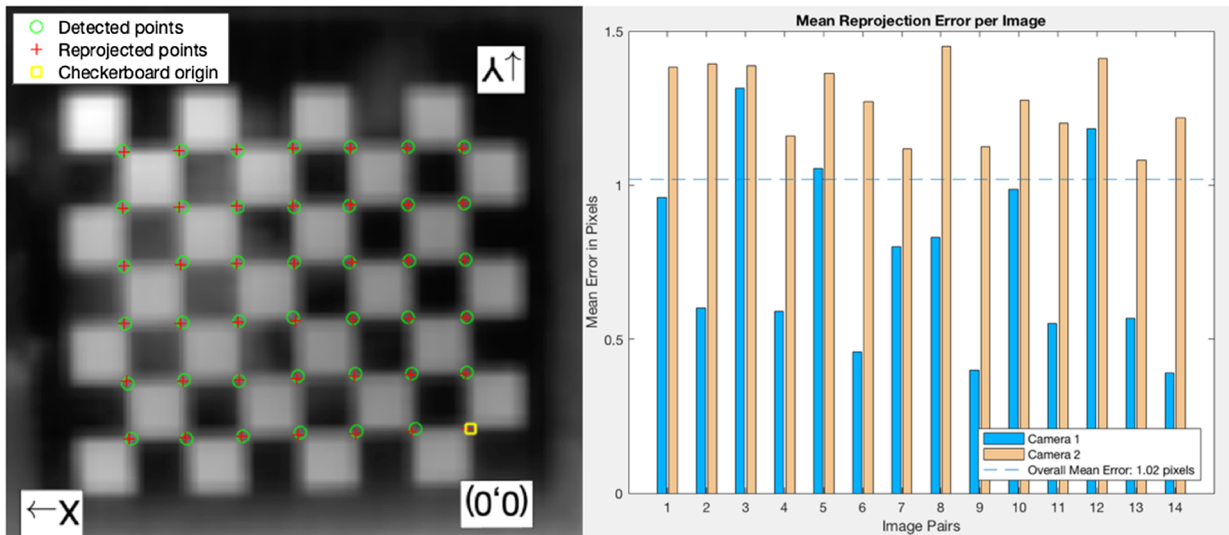


Fig. 6. Dual camera registration results: (a) detected corner points and the re-projected points after registration; (b) mean re-projection error in pixels.

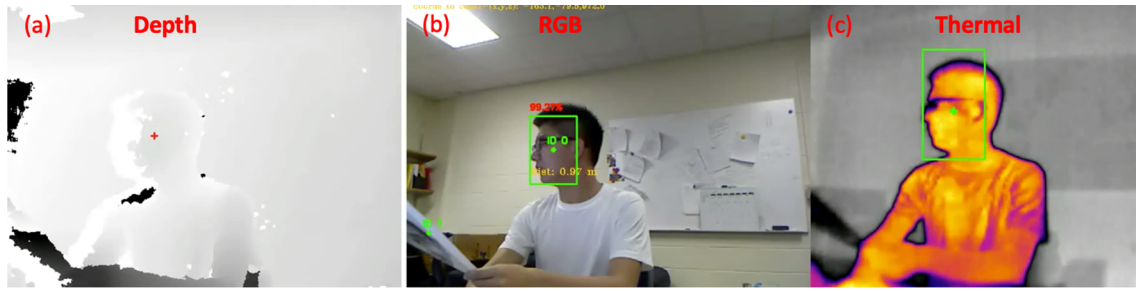


Fig. 7. Dual camera face detection: (a) depth data from the Kinect; (b) RGB images from the Kinect (for face detection); (c) thermal images from the thermal camera (bounding box is mapped from the RGB image).

camera network can be recursively configured to maximize the aggregate observability of facial regions of occupants, which is an optimization problem that is subject to a series of constraints such as the number of camera nodes, and distances between cameras and occupants. However, this is beyond the scope of this study.

3.2.2. Occupant registration among different camera nodes in the network

In Section 3.1.2, the authors introduced the centroid tracking algorithm (which was implemented in a single camera node) to track multiple occupants across video frames. However, in a camera-occupant network, thermal profiles collected by multiple camera nodes from different viewpoints should be associated with the same occupants in the 3D world coordinate system. This process is called occupant registration. For example, as shown in Fig. 9, camera node 1 observes only two of three occupants (occupant 3 is outside of the view), while camera node 2 observes all three of these three occupants. In this case, the network should correctly associate the two thermal profiles in camera node 1 with the two corresponding occupants in camera node 2. Typically, the registration can be achieved by calculating the descriptors of feature points of occupants in different viewpoints and then

mapping these feature points according to their similarities [48]. However, this feature-based method is not suitable for the camera-occupant network as (1) feature points detected from two viewpoints in the network can be very different, e.g., one camera node observes the frontal face while another observes a profile face; and (2) calculating the feature descriptors can be computationally expensive, which is not optimal for real-time registration.

Therefore, the authors implemented the location-based occupant registration using the stereo vision and the pinhole camera model introduced in Section 3.1.3. In this case, instead of registering a thermal camera with an RGB camera (see Fig. 4), each pair of RGB cameras are registered through the same stereo vision calibration process to get the transformation matrix $[R|T]$ using a paper printed checkerboard pattern. Then, the 3D world coordinate $[X \ Y \ Z \ 1]^T$ of each occupant with respect to the world origin can be calculated from Eq. (2). Finally, the occupant ID in the camera node i can be mapped to that from a different viewpoint j based on the closest distance using Eq. (3), which is a modified version of Eq. (1).

$$\text{Occupant ID} = \underset{m \in M_j}{\operatorname{argmin}} \|x_i - m\| \quad (3)$$

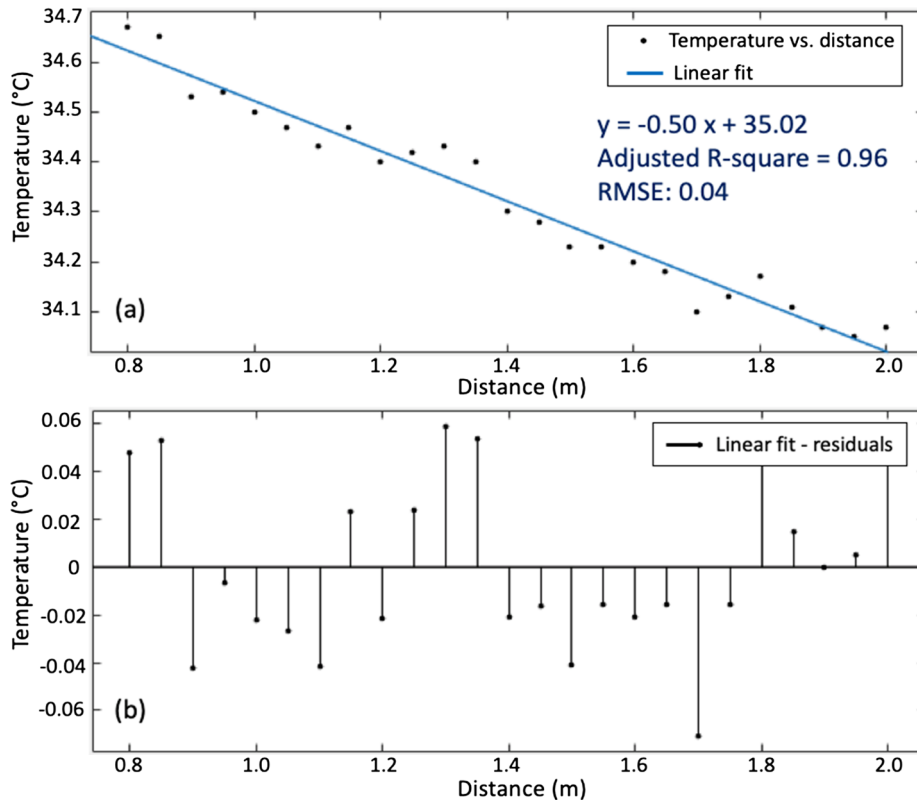


Fig. 8. A linear fit of distance and temperature measurements: (a) linear regression line; (b) linear fit residuals.

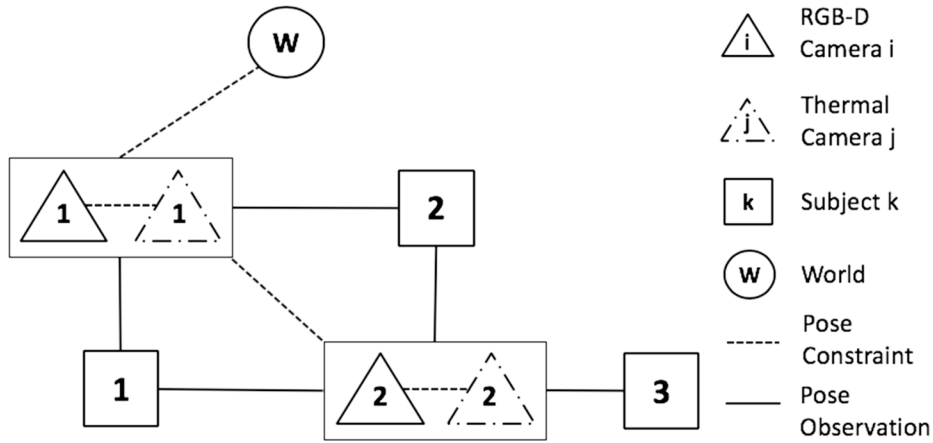


Fig. 9. Camera-occupant network.

where M_j is a set of world coordinates of all subjects in the camera node j ; m is the world coordinate of a subject in the set M_j ; x_i is the world coordinate of a subject in the camera node i (which needs to be registered); $\|\cdot\|$ is the $L2$ -norm. Thus, the camera network can recursively register all occupants observed by different camera nodes.

3.2.3. Data communication among camera nodes

In the camera network, camera nodes need to exchange occupants' world coordinates and register the same occupant from different viewpoints. For the scalability purpose, the program was coded for each single camera node such that the network can be quickly configured when adding or removing camera nodes. In this study, the data communication was implemented using the User Datagram Protocol (UDP). UDP has advantages of low latency and loss-tolerating, which are suitable for real-time video streaming.

3.3. Data cleaning and feature extraction

The facial skin temperature collected directly from each bounding box in thermal images are the raw data which can contain several types of random noises such as the false detection of background as faces, inaccurate face coordinates mapping due to occlusions, and interference of a high temperature object in the environment (e.g., hot water cup). These noises are typically shown as out-of-range isolated noises in the measurements. As a result, the authors applied the median filter shown in Eq. (4) to remove such noises before data analysis.

$$y[k] = \text{median}\{x[i], i \in w\} \quad (4)$$

where $y[k]$ is the k th value after filtering; w is a neighborhood defined by the user; $x[i]$ is the raw data in the neighborhood w .

Then, the moving average filter as shown in Eq. (5) was applied to further filter out noises from fluctuations.

$$y[k] = \frac{1}{2n+1} \sum_{i=-n}^{i=n} y[k+i] \quad (5)$$

where $y[k]$ is the k th value after filtering; $2n+1$ is the window size of the moving average; $y[k+i]$ is the raw data in the sliding window.

As images of frontal faces are not guaranteed in the camera network, unlike Li et al. [31] which segmented the frontal face into six local facial regions (e.g., forehead, nose) and extracted skin temperature from each local region as the features of personal comfort models, this study extracted the skin temperature from the whole facial region which consists of both frontal and profile faces. The features collected from the detected facial region are summarized as follows:

- The mean, first quartile, third quartile, and maximum of all pixels in the detected facial region. These features describe the distribution of

skin temperature over a facial region.

- The skin temperature variance of all pixels in the detected facial region. As suggested in Li et al. [31], the nose, ears, and cheeks regions have larger skin temperature variations than other regions when the ambient air temperature changes. Therefore, large skin temperature variations over the whole facial region can imply that an individual is experiencing cold stress (as some local regions become significantly colder than others).
- The skin temperature gradients of every minute. As suggested in Li et al. [31], the gradients can imply the heat or cold stress in the environment which is useful to predict an occupant's thermal comfort state.

4. Experimental setup and protocol

The proposed camera network was experimentally tested in a transient heating environment to verify its applicability in real operational built environments. The experiment included a 20-minute preparation phase and a 50-minute data collection phase to collect subjects' facial skin temperature. The experiment was conducted in a research office at the University of Michigan (UM) during the heating season in 2018. The experiment office has equipped a thermostat which can control the indoor temperature from low (23 °C) to high (27 °C) through two HVAC diffusers. As shown in Fig. 10, two dual camera nodes were placed approximately 1.3 m away from a table where subjects sat at during the experiment to represent a simplified camera network. Two temperature and humidity sensors (humidity accuracy: $\pm 5\%$, temperature accuracy: $\pm 1^\circ\text{C}$) continuously monitored the ambient environmental conditions. The two sensors were placed at the waist level (0.65 m above the floor) which was close to the specified

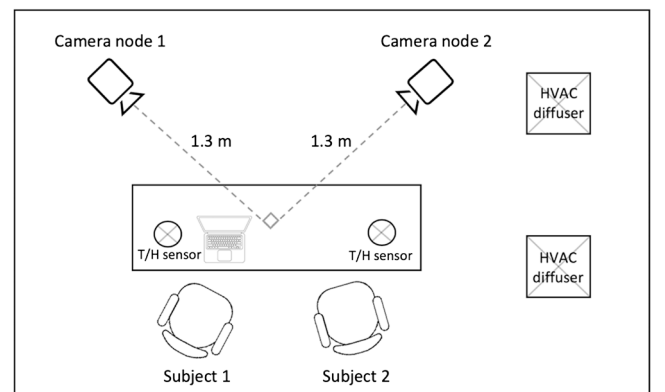


Fig. 10. The experimental setup.

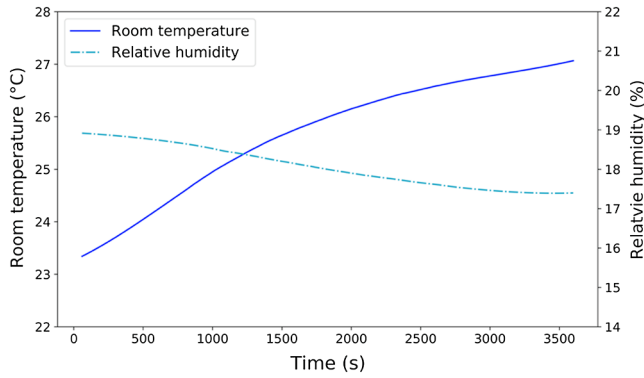


Fig. 11. Room condition in a transient heating experiment.

height of 0.6 m for seated occupants in ASHRAE standards 55. To represent a multi-occupancy scenario, 2 subjects were required to participate in the experiment each time. In total, 16 subjects (10 males and 6 females for a total of 8 experiments) were recruited. All subjects were UM students and were healthy at the time of the experiment. The experiment has been approved by the UM Institutional Review Board for conducting human subjects research.

Before the experiment started, the room temperature was set at 23 °C to represent a cool environment. During the 20-minute preparation phase, subjects were asked to remain seated in the testbed to reach a steady-state skin temperature. Then, during the following 50-minute data collection phase, the thermostat was set at 27 °C to create a transient heating environment (see Fig. 11). During this period, subjects were asked to perform daily office activities such as reading, typing, browsing, or chatting with each other while their facial skin temperature was extracted by the camera network. To collect the ground truth thermal comfort, subjects were required to report their thermal sensations in a five-point scale (from “cold” to “hot”) and preferences in a three-point scale (from “prefer warmer” to “prefer cooler”) through a phone application and also wear a wristband sensor (Microsoft Band 2) to record the wrist skin temperature (only for the purpose of comparison). For more details about the phone application and the wristband, please refer to Li et al. [9]. After the experiment, subjects participated in a survey to evaluate their experience regarding the user acceptance, privacy concern, and level of intrusiveness of the camera network.

It is worth noting that unlike existing experimental studies which typically require subjects to stay in the same posture and refrain from body movements during data collection (e.g., [30,33]), subjects in this study were allowed to move their body (e.g., stretching), change their postures and facing directions, or even move around in the room to represent scenarios in real office settings and also make them feel as comfortable as possible (to achieve the least intrusiveness caused by the system). The authors believe such an experimental study verifies the applicability of the proposed system.

Fig. 12 shows the image frames collected in the experiment. The two rows of images are the views of camera nodes 1 and 2, respectively. The three columns are the depth images, RGB images, and thermal images collected from the dual camera nodes. It can be seen that for the right subject (denoted in a green bounding box), camera node 1 detects the profile face at a distance of 1.11 m while camera node 2 detects the frontal face at a distance of 1.30 m. This demonstrates the idea of a camera network which observes each subject from different angles and distances to overcome the limitations of a single camera. Skin temperature features (discussed in Section 3.3) are automatically and continuously extracted from the identified facial region and compensated by viewing distances (discussed in Section 3.1.4) for data analysis.

5. Results and discussion

In this section, we presented the statistics of facial skin temperature features (discussed in Section 3.3) and their correlations during the thermoregulation process. Correlations between the facial mean skin temperature and wrist skin temperature were also evaluated to validate the proposed networked camera system. Then, we mapped facial mean skin temperature to each subject's feedback and explored if an individual's thermal comfort state can be reflected by the facial skin temperature. Finally, a post-experiment survey was conducted to evaluate subjects' experience with the proposed approach.

5.1. Summary of facial skin temperature features and gender differences

Tables 5 and 6 present a summary of skin temperature features collected by camera node 1 and 2, correspondingly. The mean ($\bar{\mu}$) and standard deviation ($\bar{s}$) of each feature are calculated using Eq. (6) and (7) as shown below.

$$\bar{\mu} = \frac{1}{n} \sum_{k=1}^n \mu_k, \quad SD(\bar{\mu}) = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (\mu_k - \bar{\mu})^2} \quad (6)$$

$$\bar{s} = \frac{1}{n} \sum_{k=1}^n s_k, \quad SD(\bar{s}) = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (s_k - \bar{s})^2} \quad (7)$$

where $\bar{\mu}$ and $SD(\bar{\mu})$ are the mean and standard deviation of skin temperature features; $\bar{s}$ and $SD(\bar{s})$ are the mean and standard deviation of the sample standard deviation; n is the number of subjects which is 16 in this study.

Statistics in Tables 5 and 6 are divided into three categories: all subjects, all males, and all females, respectively. For each category, two statistics $\bar{\mu} \pm SD(\bar{\mu})$ and $\bar{s} \pm SD(\bar{s})$ are calculated for each feature: $\bar{\mu} \pm SD(\bar{\mu})$ indicates the mean value of each feature and its variations across different subjects, and $\bar{s} \pm SD(\bar{s})$ reflects how much each feature changes in the experiment and its variations across subjects. As shown in Tables 5 and 6, the first quartile temperature of the facial region has a higher variation (node 1: 0.72 ± 0.16 °C, node 2: 0.63 ± 0.20 °C) than all other features while the maximum temperature feature has a smaller variation (node 1: 0.43 ± 0.10 °C, node 2: 0.41 ± 0.10 °C). This result suggests that as room temperature gradually increases in the experiment, a subject's facial skin temperature increases correspondingly due to the thermoregulation, however, gaps between the baseline temperature and high temperature regions tend to become narrower over time. This finding is also partially reflected by the facial temperature variances (node 1: 0.40 ± 0.16 °C, node 2: 0.59 ± 0.26 °C) as 9 out of 16 subjects observe a slight decrease in the variances over time, which is supported by Li et al. [31] where low skin temperature regions such as nose and cheeks change more significantly than high temperature regions like forehead under heat stress. For the rest of subjects who hold steady or increasing variances, this may be caused by the frequent changes of their facing directions and body movements such that a camera node does not observe the same facial region over time.

As shown in Tables 5 and 6, females generally have a slightly higher skin temperature and larger variations than males. To evaluate if there exist significant gender differences in each feature, the t -test is conducted (two-sided 95% confidence interval). The result suggests that except one group (denoted in the light grey shading in Table 6), males and females do not show significant differences in the selected skin temperature features.

By comparing Tables 5 and 6, it can be found that the facial skin temperature ($\bar{\mu}$) and variation ($\bar{s}$) of each feature collected by the camera node 1 are slightly higher than those by the camera node 2. This



Fig. 12. Image frames collected in the experiment (row 1: views from camera node 1; row 2: views from camera node 2).

Table 5

Statistics of skin temperature features collected by camera node 1.

Features	All subjects		Males		Females	
	$\bar{\mu} \pm SD(\bar{\mu})$	$\bar{s} \pm SD(\bar{s})$	$\bar{\mu} \pm SD(\bar{\mu})$	$\bar{s} \pm SD(\bar{s})$	$\bar{\mu} \pm SD(\bar{\mu})$	$\bar{s} \pm SD(\bar{s})$
Mean	32.76 $\pm$ 0.38	0.55 $\pm$ 0.12	32.64 $\pm$ 0.38	0.51 $\pm$ 0.12	32.98 $\pm$ 0.31	0.60 $\pm$ 0.12
1st quartile	32.12 $\pm$ 0.46	0.72 $\pm$ 0.16	31.99 $\pm$ 0.45	0.66 $\pm$ 0.16	32.35 $\pm$ 0.42	0.82 $\pm$ 0.10
3rd quartile	33.86 $\pm$ 0.40	0.55 $\pm$ 0.15	33.75 $\pm$ 0.43	0.53 $\pm$ 0.15	34.04 $\pm$ 0.29	0.58 $\pm$ 0.15
Max	35.02 $\pm$ 0.31	0.43 $\pm$ 0.10	34.93 $\pm$ 0.31	0.43 $\pm$ 0.11	35.18 $\pm$ 0.27	0.44 $\pm$ 0.11
Variance	2.53 $\pm$ 0.32	0.40 $\pm$ 0.16	2.54 $\pm$ 0.38	0.39 $\pm$ 0.16	2.51 $\pm$ 0.23	0.41 $\pm$ 0.19

Note: All numbers are in $^{\circ}\text{C}$.

Table 6

Statistics of skin temperature features collected by camera node 2.

Features	All subjects		Males		Females	
	$\bar{\mu} \pm SD(\bar{\mu})$	$\bar{s} \pm SD(\bar{s})$	$\bar{\mu} \pm SD(\bar{\mu})$	$\bar{s} \pm SD(\bar{s})$	$\bar{\mu} \pm SD(\bar{\mu})$	$\bar{s} \pm SD(\bar{s})$
Mean	32.16 $\pm$ 0.22	0.47 $\pm$ 0.16	32.08 $\pm$ 0.21	0.45 $\pm$ 0.15	32.29 $\pm$ 0.16	0.49 $\pm$ 0.18
1st quartile	31.44 $\pm$ 0.32	0.63 $\pm$ 0.20	31.32 $\pm$ 0.29	0.60 $\pm$ 0.19	31.65 $\pm$ 0.27	0.68 $\pm$ 0.22
3rd quartile	33.17 $\pm$ 0.31	0.48 $\pm$ 0.13	33.12 $\pm$ 0.35	0.47 $\pm$ 0.13	33.26 $\pm$ 0.22	0.50 $\pm$ 0.15
Max	34.32 $\pm$ 0.27	0.41 $\pm$ 0.10	34.29 $\pm$ 0.33	0.40 $\pm$ 0.09	34.37 $\pm$ 0.14	0.42 $\pm$ 0.13
Variance	2.24 $\pm$ 0.79	0.59 $\pm$ 0.26	2.26 $\pm$ 0.82	0.58 $\pm$ 0.29	2.20 $\pm$ 0.81	0.62 $\pm$ 0.22

Note: All numbers are in $^{\circ}\text{C}$. Grey shading indicates the means of two groups are statistically different.

difference can be caused by the fact that one camera node captures more warmer regions than the other in the experiment (i.e., node 1 observes more frames of the frontal face which contains the forehead region that is typically warmer than cheeks). However, the trends of skin temperature features resulted from the two camera nodes are both increasing over time, which can be used to interpret an individual's comfort state. This will be further discussed in Sections 5.2 and 5.3.

5.2. Correlation analysis between different skin temperature features

Correlation analysis is conducted to further investigate the relationships between different features. As shown in Tables 7 and 8, the Pearson correlation coefficients of all features except for variances suggest a strong positive correlation between each feature pair (ranged from 0.63 to 0.94). The weak correlations between variances and other features indicate that variances are relatively steady in the experiment compared to others. To reduce the dimensionality of features, we adopt the facial mean skin temperature of the whole facial region detected in

Table 7

Correlations between skin temperature features collected by camera node 1.

	Mean	1st quartile	3rd quartile	Max	Var
Mean	1.00	0.92	0.91	0.88	-0.00
1st quartile	0.92	1.00	0.78	0.75	-0.16
3rd quartile	0.91	0.78	1.00	0.94	0.15
Max	0.88	0.75	0.94	1.00	0.19
Var	-0.00	-0.16	0.15	0.19	1.00

Table 8

Correlations between skin temperature features collected by camera node 2.

	Mean	1st quartile	3rd quartile	Max	Var
Mean	1.00	0.88	0.83	0.82	-0.11
1st quartile	0.88	1.00	0.63	0.63	-0.35
3rd quartile	0.83	0.63	1.00	0.94	0.13
Max	0.82	0.63	0.94	1.00	0.20
Var	-0.11	-0.35	0.13	0.20	1.00

Table 9

Pearson correlation coefficients between the facial mean skin temperature and wrist skin temperature of each subject.

Subject ID	1	2	3	4	5	6	7	8
Node 1	0.78	0.74	0.84	0.71	0.77	0.80	0.62	0.62
Node 2	0.77	0.77	0.73	0.71	0.75	0.79	0.60	0.64
Subject ID	9	10	11	12	13	14	15	16
Node 1	0.47	0.82	0.70	0.64	0.84	0.92	0.77	0.77
Node 2	0.50	0.52	0.49	0.45	0.65	0.64	0.59	-0.21

the cameras as the main feature for further analysis because it is representative of other features (due to the high correlations) and also more precise than others due to averaging.

5.3. Correlation analysis between the facial mean skin temperature and wrist skin temperature

Skin temperature from a wristband sensor is used to validate the proposed camera network. The wristband sensor has a resolution of 1 °C and takes a measurement every 30 s. Due to the differences between the wrist and facial skin temperature and the systematic error of different instruments, measurements from these two sources are not directly compared. Instead, we analyze their correlations as room temperature increases. As shown in Table 9, all coefficients, except one (subject 16, node 2), suggest moderate to strong positive correlations between the facial mean skin temperature and wrist temperature (ranged from 0.45 to 0.92), which indicates the non-intrusive camera network is able to capture the same thermoregulatory responses as the wearables. Also, for the same subject, it can be seen that the coefficient resulted from one camera can be much higher than that of the other (e.g., subject 10). This finding suggests that despite an occupant can be observed by

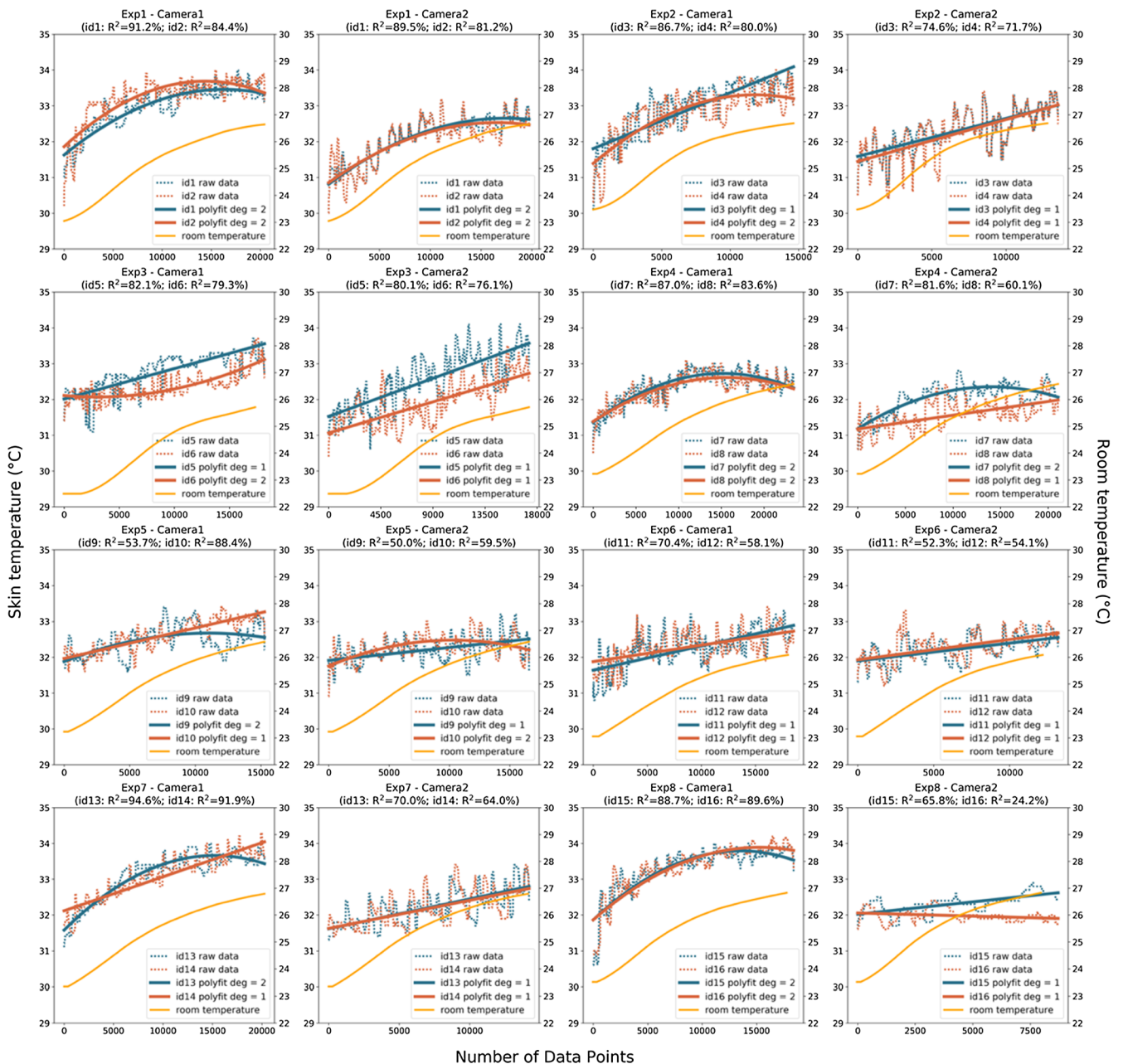


Fig. 13. Polynomial fit of the facial mean skin temperature of each subject.

Table 10
t-tests between the starting and ending facial mean temperature in the experiment.

Subject	Camera node	$\bar{T}_{start}$ (°C)	$\bar{T}_{end}$ (°C)	p-value	TCS_{start}	TCS_{end}
Exp. 1	Id1	1	31.6	< 0.001	Cold (warmer)	Warm (no change)
		2	31.2	< 0.001		
	Id2	1	31.5	< 0.001	Cold (warmer)	Hot (cooler)
		2	31.4	< 0.001		
Exp. 2	Id3	1	31.8	< 0.001	Cold (warmer)	Warm (no change)
		2	31.6	< 0.001		
	Id4	1	31.4	< 0.001	Cold (warmer)	Neutral (no change)
		2	31.7	< 0.001		
Exp. 3	Id5	1	32.1	< 0.001	Cool (warmer)	Warm (cooler)
		2	31.5	< 0.001		
	Id6	1	32.1	< 0.001	Cool (warmer)	Neutral (no change)
		2	31.3	< 0.001		
Exp. 4	Id7	1	31.3	< 0.001	Cool (no change)	Warm (cooler)
		2	31.3	< 0.001		
	Id8	1	31.4	< 0.001	Cold (warmer)	Neutral (no change)
		2	31.1	< 0.001		
Exp. 5	Id9	1	32.1	< 0.001	Cool (warmer)	Neutral (no change)
		2	32.0	< 0.001		
	Id10	1	32.1	< 0.001	Neutral (no change)	Warm (no change)
		2	31.9	< 0.001		
Exp. 6	Id11	1	31.4	< 0.001	Cool (warmer)	Neutral (no change)
		2	32.0	< 0.001		
	Id12	1	31.6	< 0.001	Neutral (no change)	Warm (no change)
		2	32.0	< 0.001		
Exp. 7	Id13	1	31.7	< 0.001	Cold (warmer)	Warm (no change)
		2	31.7	< 0.001		
	Id14	1	31.9	< 0.001	Neutral (no change)	Neutral (cooler)
		2	31.7	< 0.001		
Exp. 8	Id15	1	31.7	< 0.001	Cool (no change)	Warm (cooler)
		2	32.2	< 0.001		
	Id16	1	32.1	< 0.001	Neutral (no change)	Hot (cooler)
		2	32.1	< 0.001		

Note: Subjects' thermal preferences are shown in the parentheses in the TCS columns.

multiple camera nodes, data from different cameras may not be equally important in assessing changes in their skin temperature. Therefore, future studies by the authors will assign different weights to cameras in the comfort assessment according to their viewing distances and angles.

5.4. Mapping facial mean skin temperature to thermal comfort state

To visualize the changes of skin temperature in the transient heating experiment, each subject's facial mean skin temperature is fitted using the polynomial regression. Specifically, the lower degree of polynomials is selected if the coefficient of determination R^2 does not increase significantly with a higher degree of polynomials (we used 5% as a threshold). As shown in Fig. 13, each subplot shows the result of an experiment measured by a camera node. In each subplot, the two dotted lines denote the processed skin temperature data of two subjects (discussed in Section 3.3), which are then fitted by polynomials of degree ranged from 1 to 3. All fitted curves (denoted in solid lines of the corresponding colors) in Fig. 13, except subject 16, demonstrate an increasing trend of skin temperature over time, which implies the increases in skin blood flow under heat stress.

To map the thermoregulatory process to thermal comfort state, facial mean skin temperature of the first five minutes (denoted as $\bar{T}_{start}$) and the last five minutes (denoted as $\bar{T}_{end}$) in the heating experiment (from 23 °C to 27 °C) are calculated to represent a subject's starting and ending physiological states, respectively. As shown in Table 10, two-tailed t-test shows that all 16 subjects have a statistically higher facial skin temperature ($p < .001$) as the room temperature increases in the experiment. In addition, all subjects reported distinct thermal comfort states (i.e., thermal sensation and preference) at these two stages, which are denoted as TCS_{start} and TCS_{end} in Table 10. As an example, Fig. 14 presents a subject's thermal sensation votes and his/her corresponding facial mean skin temperature. Results of the analysis of variance (ANOVA) show that the means of different votes are significantly different, which suggests that facial skin temperature collected by the proposed camera network can serve as an indicator of subjects' thermal comfort state.

5.5. Post-experiment evaluations of user experience

A survey is distributed to all subjects to understand their experience

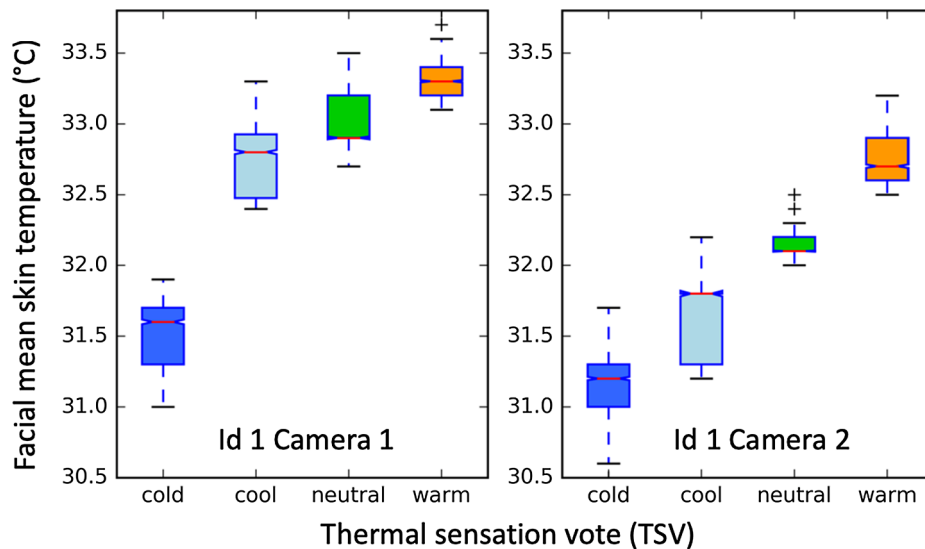


Fig. 14. Thermal sensation vote and the corresponding facial mean skin temperature of a subject.

Please rate the degree to which you agree or disagree with each of the following statements related to the camera network (5 - Strongly agree, 4 - Agree, 3 - Neutral, 2 - Disagree, 1 - Strongly disagree):

User Acceptance – I would recommend such camera network systems to be applied in multi-occupancy spaces to assess occupants' thermal comfort (e.g., classroom, rest lounge).

Privacy Concern – I do NOT have any privacy concern about such camera network systems if they are applied in the built environment.

Level of Intrusiveness Q1 – Based on my experience in the experiment, I think the camera network is not intrusive at all.

Level of Intrusiveness Q2 – Compared to wristband sensors, I prefer using the camera network to collect my skin temperature data as it is less interruptive and does not cause any pain or strain that may arise from wearables.

Fig. 15. Post-experiment survey questions.

Table 11
Subjects' post-experiment evaluation.

Subject ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Mean
User Acceptance	4	4	4	5	4	5	5	4	4	3	4	5	5	4	4	5	4.31
Privacy Concern	4	4	4	3	4	3	5	4	4	4	4	4	4	4	4	4	3.94
Intrusiveness Q1	4	4	4	3	5	5	5	4	5	4	5	5	5	5	4	5	4.50
Intrusiveness Q2	5	5	5	4	5	5	5	4	5	5	5	5	5	5	4	5	4.81

with the camera network regarding the user acceptance, privacy concerns, and level of intrusiveness caused by the system (see Fig. 15). The feedback suggests that the proposed approach has a high user acceptance regarding its application in the built environment (mean: 4.31, see Table 11), relatively low privacy concern caused by the use of cameras (mean: 3.94), and strong agreement in the non-intrusiveness of the approach (Q1 mean: 4.50, Q2 mean: 4.81).

5.6. Limitations and future work

Four limitations of this study should be acknowledged. First, the skin temperature collected from frontal and profile faces is not differentiated in the current approach, which can be a reason for the fluctuations in the measurements. As a result, future studies by the authors will keep the frontal and profile faces (or possibly different facing

directions) as separate datasets to evaluate subjects' thermal comfort state. In addition, previous studies by the authors suggested that regions such as cheeks and noses are more indicative of an individual's thermoregulatory process. Thus, if these regions are detected, a larger weight can be assigned to this frame in the comfort assessment.

Second, the proposed camera network is tested in a simplified multi-occupancy environment with two subjects. Several challenges may arise from a larger space with more subjects, such as occlusions, increased viewing distance, and occupant registration. Thus, the scalability potential of the proposed method requires further investigation.

Third, this study aims to address the fundamental questions related to the proposed non-intrusive camera network, such as its components, camera registration, and feature extraction, as well as demonstrate the correlations between facial skin temperature and subjects' thermal comfort feedback. Thus, future work can focus on developing personal

comfort models using machine learning methods and the weighted skin temperature data from different camera nodes to continuously predict each subject's thermal comfort state.

Fourth, despite the low-cost thermal camera can achieve an acceptable accuracy as suggested by Li et al. [31], proper calibration of the thermal camera with a contact thermometer may improve the predictive power of personal comfort models.

6. Conclusions

This study introduced the main characteristics of the non-intrusive detection of thermal comfort and proposed a low-cost networked camera system to non-intrusively measure occupants' facial skin temperature for real-time thermal comfort assessment in multi-occupancy environments. Each camera node in the network fuses the RGB-D and thermal images collected from a Kinect and a low-cost thermal camera. The experimental results from 16 subjects suggest that the variations in low temperature facial regions are more significant than high temperature regions under heat stress, as well as moderate to strong positive correlations between the skin temperature collected by the camera network and wearables (ranged from 0.45 to 0.92). Moreover, subjects' facial skin temperature has observed statistically significant increases when the room temperature changes from 23 °C to 27 °C. Results of ANOVA support our assumption that facial skin temperature can serve as an indicator of one's thermal comfort state. Finally, subjects have expressed positive evaluation regarding the usefulness, privacy issues, and the non-intrusiveness of the proposed approach.

This pilot study has the potential to transition the current human physiological sensing from an intrusive and wearable device-based approach to a truly non-intrusive and scalable approach such that skin temperature can be automatically measured without any constraints on occupants' activities or participation. The proposed camera network can be incorporated into the building HVAC systems for energy control and thermal comfort management. For example, a real-time interpretation of thermal comfort allows the HVAC systems to dynamically adjust its setpoint and air flow and also select the optimum settings to maximize the overall comfort. If the indoor environment is equipped with personal heating/cooling devices or HVAC zoning systems, personalized conditioning can be delivered to the corresponding location if conflicts of thermal preference exist in the shared space. The proposed approach is particularly promising in multi-occupancy environments, such as offices, conference rooms, rest lounges as personal and wearable devices may not be available for everyone. Knowledge gained from this study can also be applied to other critical built environments, including the transportation systems, health facilities, and extreme working environments where occupants' thermal comfort and satisfaction are much needed. Future work by the authors will deploy a functional prototype in a testbed and connect it with a smart thermostat to investigate if comfort can be improved while enhancing the building's energy performance.

Acknowledgments

The authors would like to acknowledge the financial support for this research received from the U.S. National Science Foundation (NSF) CBET 1349921 and CBET 1804321. Any opinions and findings in this paper are those of the authors and do not necessarily represent those of the NSF. The authors would like to thank all subjects who participated in the experimental studies.

References

- [1] European Commission. Buildings; 2018. <https://ec.europa.eu/energy/en/topics/energy-efficiency/buildings>.
- [2] Energy Information Administration (EIA). How much energy is consumed in residential and commercial buildings in the United States? <https://www.eia.gov/tools/faqs/faq.php?id=86&t=1>.
- [3] Department of Energy (DOE). Heating and cooling; 2017. <https://energy.gov/public-services/homes/heating-cooling>.
- [4] Karmann C, Schiavon S, Arens E. Percentage of commercial buildings showing at least 80% occupant satisfied with their thermal comfort; 2018.
- [5] Roulet CA, Johnner N, Foradini F, Bluyssen P, Cox C, de Oliveira Fernandes E, et al. Perceived health and comfort in relation to energy use and building characteristics. *Build Res Inf* 2006;34(5):467–74.
- [6] Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Exposure Sci Environ Epidemiol* 2001;11(3):231.
- [7] Fang L, Wyon DP, Clausen G, Fanger PO. Impact of indoor air temperature and humidity in an office on perceived air quality, SBS symptoms and performance. *Indoor Air* 2004;14:74–81.
- [8] ASHRAE. ANSI/ASHRAE Standard 55-2010. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.; 2010.
- [9] Li D, Menassa CC, Kamat VR. Personalized human comfort in indoor building environments under diverse conditioning modes. *Build Environ* 2017;126:304–17.
- [10] Ning H, Wang Z, Ji Y. Thermal history and adaptation: Does a long-term indoor thermal exposure impact human thermal adaptability? *Appl Energy* 2016;183:22–30.
- [11] Parsons K. Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort, and performance. CRC Press; 2014.
- [12] Ugursal A, Culp CH. The effect of temperature, metabolic rate and dynamic localized airflow on thermal comfort. *Appl Energy* 2013;111:64–73.
- [13] Fanger PO. Thermal comfort. Analysis and applications in environmental engineering. Thermal comfort. Analysis and applications in environmental engineering; 1970.
- [14] Brager G, de Dear R. A standard for natural ventilation; 2000.
- [15] Chaudhuri T, Zhai D, Soh YC, Li H, Xie L. Thermal comfort prediction using normalized skin temperature in a uniform built environment. *Energy Build* 2018;159:426–40.
- [16] Jung W, Jazizadeh F. Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions. *Appl Energy* 2019;239:1471–508.
- [17] Kim J, Zhou Y, Schiavon S, Raftery P, Brager G. Personal comfort models: predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Build Environ* 2018;129:96–106.
- [18] Li D, Menassa CC, Kamat VR. A personalized HVAC control smartphone application framework for improved human health and well-being. *Computing in civil engineering* 2017. p. 82–90.
- [19] Allab Y, Pellegrino M, Guo X, Nefzaoui E, Kindin A. Energy and comfort assessment in educational building: Case study in a French university campus. *Energy Build* 2017;143:202–19.
- [20] Bourdeau M, Guo X, Nefzaoui E. Buildings energy consumption generation gap: A post-occupancy assessment in a case study of three higher education buildings. *Energy Build* 2018;159:600–11.
- [21] Li D, Menassa CC, Karatas A. Energy use behaviors in buildings: towards an integrated conceptual framework. *Energy Res Social Sci* 2017;23:97–112.
- [22] Thomas A, Menassa CC, Kamat VR. Lightweight and adaptive building simulation (LABS) framework for integrated building energy and thermal comfort analysis. *Building simulation*, vol. 10, no. 6. Tsinghua University Press; 2017. p. 1023–44.
- [23] Feldmeier M, Paradiso JA. Personalized HVAC control system. *Internet of Things (IoT)*. IEEE; 2010. p. 1–8.
- [24] Ghahramani A, Castro G, Karvigh SA, Becerik-Gerber B. Towards unsupervised learning of thermal comfort using infrared thermography. *Appl Energy* 2018;211:41–9.
- [25] Shahzad S, Calautit JK, Aquino AI, Nasir DS, Hughes BR. A user-controlled thermal chair for an open plan workplace: CFD and field studies of thermal comfort performance. *Appl Energy* 2017;207:283–93.
- [26] Jazizadeh F, Jung W. Personalized thermal comfort inference using RGB video images for distributed HVAC control. *Appl Energy* 2018;220:829–41.
- [27] Jazizadeh F, Ghahramani A, Becerik-Gerber B, Kichkaylo T, Orosz M. Human-building interaction framework for personalized thermal comfort-driven systems in office buildings. *J Comput Civil Eng* 2013;28(1):2–16.
- [28] FLIR. Temperature guns versus thermal imaging technology; 2015. <https://www.flir.co.uk/discover/rd-science/temperature-guns-versus-thermal-imaging-technology/> [November 05, 2015].
- [29] Ranjan J, Scott J. ThermalSense: determining dynamic thermal comfort preferences using thermographic imaging. *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. ACM; 2016. p. 1212–22.
- [30] Metzmacher H, Wölki D, Schmidt C, Frisch J, van Treeck C. Real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment. *Energy Build* 2018;158:1063–78.
- [31] Li D, Menassa CC, Kamat VR. Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography. *Energy Build* 2018;176:246–61.
- [32] Kwon S, Kim H, Park KS. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. *Annual international conference of the IEEE on engineering in medicine and biology society (EMBC)*. IEEE; 2012. p. 2174–7.
- [33] Jung W, Jazizadeh F. Vision-based thermal comfort quantification for HVAC control. *Build Environ* 2018;142:513–23.
- [34] Jung W, Jazizadeh F. Towards integration of doppler radar sensors into personalized thermoregulation-based control of HVAC. *Proceedings of the 4th ACM international conference on systems for energy-efficient built environments*. ACM; 2017.

- p. 21.
- [35] Thorne CH. Grabb and Smith's plastic surgery. Lippincott Williams & Wilkins; 2013.
 - [36] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. *Null. IEEE*; 2001. p. 511.
 - [37] Dalal N, Triggs B. Histograms of oriented gradients for human detection. *International conference on computer vision & pattern recognition (CVPR'05)*, vol. 1. IEEE Computer Society; 2005. p. 886–93.
 - [38] Turk MA, Pentland AP. Face recognition using eigenfaces. *Proceedings 1991 IEEE computer society conference on computer vision and pattern recognition. IEEE*; 1991. p. 586–91.
 - [39] Jiang H, Learned-Miller E. Face detection with the faster R-CNN. *12th IEEE international conference on automatic face & gesture recognition (FG 2017)*. IEEE; 2017. p. 650–7.
 - [40] Rosebrock A. Simple object tracking with OpenCV; 2018. <https://www.pyimagesearch.com/2018/07/23/simple-object-tracking-with-opencv/>.
 - [41] OpenCV. Camera calibration and 3D reconstruction; 2018. https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html.
 - [42] OpenCV 3.3. Deep neural networks; 2018. https://docs.opencv.org/3.4/d2/d58/tutorial_table_of_content_dnn.html.
 - [43] Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. *The proceedings of the seventh IEEE international conference on computer vision*, vol. 1. IEEE; 1999. p. 666–73.
 - [44] Mathworks. Stereo camera calibration app; 2016. <https://www.mathworks.com/help/vision/ug/stereo-camera-calibrator-app.html>.
 - [45] FLIR. 5 factors influencing radiometric temperature measurements; 2016. https://groupgets-files.s3.amazonaws.com/lepton/Guidebook_Cores_5_Factors_Influencing_Radiometric_Temperature_Measurements_Americas_.pdf [October 11, 2017].
 - [46] Khoshelham K, Elberink SO. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors* 2012;12(2):1437–54.
 - [47] Feng C, Kamat VR, Cai H. Camera marker networks for articulated machine pose estimation. *Autom Constr* 2018;96:148–60.
 - [48] Lowe DG. Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 2004;60(2):91–110.