

Skin Temperature Extraction Using Facial Landmark Detection and Thermal Imaging for Comfort Assessment

Ashrant Aryal

University of Southern California
Los Angeles, California
aaryal@usc.edu

Burcin Becerik-Gerber

University of Southern California
Los Angeles, California
becerik@usc.edu

ABSTRACT

Despite the large share of energy consumption, current HVAC systems in buildings fail to meet their primary purpose of maintaining comfortable indoor conditions. Current “one size fits all” approach to control the thermal conditions in an environment lead to a high degree of occupant dissatisfaction. Advancements in Internet of Things and Machine Learning have opened the possibility of deploying different sensors at a wide scale to monitor environmental and physiological information and using collected sensor data to model individual comfort requirements. Thermal imaging has recently gained interest as one of the possible ways to monitor physiological information (skin temperature) for thermal comfort assessment. Previous studies have shown that skin temperatures from different regions of the face, such as forehead, nose, cheeks and ears can provide useful information for predicting thermal sensation at an individual level. However, existing approaches to process thermal images either rely on manual temperature extraction or use methods that are less reliable in accurately identifying different facial regions. One of the major challenges of using thermal imaging for monitoring skin temperatures in actual buildings is that occupants may move relative to the camera. It is not practical to expect building occupants to be oriented facing the cameras at all times, therefore, it is important to be able to extract as much information as possible from instances where it is feasible to extract relevant information. In this paper, we describe an approach to extract skin temperature by locating specific regions of the face in thermal images. The approach involves combining data from RGB images with thermal images and leveraging facial landmark detection in RGB images. We also evaluate our approach with existing approach of face detection used in previous studies. Our study demonstrates that facial landmark detection provides a more accurate calculation of different locations in the face compared to previous studies. We show an improvement in overall quantity and quality of temperature measurements extracted from thermal images compared to previous studies. More accurate temperature measurements from thermal images can improve the accuracy of thermal imaging for modeling and predicting thermal comfort.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

BuildSys '19, November 13–14, 2019, New York, NY, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-7005-9/19/11...\$15.00

<https://doi.org/10.1145/3360322.3360848>

CCS CONCEPTS

• **Applied computing** → **Engineering**; • **Human-centered computing** → *Empirical studies in HCI; Ubiquitous and mobile computing systems and tools.*

KEYWORDS

Infrared thermography, thermal imaging, thermal comfort, facial landmarks

ACM Reference Format:

Ashrant Aryal and Burcin Becerik-Gerber. 2019. Skin Temperature Extraction Using Facial Landmark Detection and Thermal Imaging for Comfort Assessment. In *The 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19)*, November 13–14, 2019, New York, NY, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3360322.3360848>

1 INTRODUCTION

Majority of energy consumed in buildings can be attributed to Heating, Ventilation and Air Conditioning (HVAC) systems. HVAC systems’ primary purpose is to maintain comfortable indoor environments, however these systems are energy intensive and consume around 43% of total building energy in the U.S. [32]. Current standards such as ASHRAE 55 [5] and ISO 7730 [15] provide guidelines on the design and evaluation of indoor environments to maintain comfortable thermal conditions. Despite consuming a large share of total building energy, current HVAC systems fail to meet the ASHRAE 55 requirement of satisfying at least 80% of the occupants [19]. A large scale survey showed that only 38% of building occupants are satisfied with their thermal environments and only 11% of buildings meet the ASHRAE requirement of satisfying 80% or more occupants [19]. The poor performance of HVAC systems in terms of occupant satisfaction results from “one size fits all” control approaches, and the inability of current HVAC systems to provide personalized control [2].

Current standards such as ASHRAE 55 [5] and ISO 7730 [15] recommend the use of the Predicted Mean Vote and Predicted Percent Dissatisfied (PMV/PPD) model for evaluating thermal conditions in air conditioned buildings. The PMV/PPD model was developed in the 1970s by P.O. Fanger based on a series of climate chamber studies that evaluated heat balance of the human body [28]. The PMV/PPD model ignores the behavioral, psychological and physiological (other than metabolic rate) factors that influence thermal comfort. A major improvement to the PMV/PPD model was suggested by de Dear and Brager in 1998 by introducing the adaptive comfort model [12]. The adaptive model considers different physiological (acclimatization), psychological (changing thermal

expectations) and behavioral (e.g., operating windows, fans) opportunities that are available to occupants to adapt in order to maintain thermal comfort. Because the PMV/PPD model is based on a heat balance model, it recommends a narrow range of temperatures as comfortable. The adaptive model considers other factors and suggests that occupants can be comfortable over a wider range of temperatures if they have different adaptive options available. The adaptive model was included in the ASHRAE 55 as a method to determine acceptable thermal conditions for naturally ventilated buildings [5, 12]. The major limitation of both the PMV/PPD and adaptive models is that they were developed based on averaged response from a large population and do not accommodate for individual differences in occupants thermal comfort requirements. Because the traditional models are based on averaged responses, the PMV model is only about 34% accurate in predicting actual thermal sensations of occupants [8].

Recent advancements in Internet of Things (IoT) and Machine Learning (ML) have opened the possibility of using distributed sensors integrated into everyday objects to collect data at a granular level and to develop individualized models [4]. To overcome the limitations of averaged comfort models, there has been a growing interest in creating comfort models that are personalized to each individual [21]. Such models can utilize data from different sensors that monitor environmental parameters such as air temperature, humidity, radiant temperature, air speed; different sensors that monitor physiological parameters such as skin temperature, heart rate, skin conductance etc. and use different ML algorithms to build personalized comfort models. The goal of such methods is to use the personal comfort models to control the HVAC system or other comfort devices (e.g. fans, heaters etc.) to improve occupant satisfaction with thermal environments [21].

Earlier approaches to building personal comfort models relied mostly on the measurement of environmental parameters such as air temperature and humidity, and utilized different ML algorithms to map the environmental measurements to individual thermal sensations [3, 11, 16]. However, some environmental parameters such as mean radiant temperature and air speed, which influence thermal comfort, are difficult to measure accurately due to lack of inexpensive and easy to use sensors. Therefore, many researchers have utilized wearable devices to monitor physiological parameters that might be more indicative of different thermal sensations. Some of the physiological parameters used in previous studies include skin temperatures from different locations, heart rate, blood pressure, skin conductance, blood oxygen saturation and so on. [7, 25, 31]. However, wearable devices may be considered intrusive, and some occupants might not be willing to wear monitoring devices. To address this issue, recent studies have focused on developing non-intrusive techniques to monitor physiological parameters. Different techniques such as thermal imaging to monitor skin temperature, Doppler radar sensors to monitor cardiovascular parameters, and photoplethysmography (PPG) using RGB cameras to monitor changes in skin blood flow have been investigated in previous studies [18, 22, 23]. Among the non-intrusive techniques, thermal imaging has been more widely studied. Results from previous studies show that thermal imaging can provide useful information towards predicting thermal sensations [10, 23, 27, 29]. However, previous studies have either manually extracted skin

temperatures from different regions of interest (ROI) or utilized automated methods that are not very accurate in extracting skin temperatures from different ROIs using a thermal camera.

In this paper, we describe a method to automatically extract skin temperatures from different ROIs in the face. The approach involves combining data from RGB images with thermal images and leveraging facial landmark detection in RGB images. We also evaluate our approach with existing approach of face detection used in previous studies. The rest of the paper is organized as follows: section 2 gives a brief overview of relevant literature, section 3 describes our process of extracting temperatures from different ROIs and section 4 describes the evaluation of our approach against previous approaches. In section 5 we discuss different aspects around real life deployment, current limitations and highlight directions for future research. We conclude the paper in section 6.

2 LITERATURE REVIEW

The thermoregulation process in humans maintains the core body temperature within a narrow range of 36°C to 38°C by regulation of skin blood flow, shivering and sweating [1]. Shivering occurs in cold environments to increase heat generation, and sweating occurs in hot environments to increase heat loss from the body. Under cool environments, the skin blood vessels constrict (vasoconstriction) which reduces skin blood flow and results in lower skin temperature. Under warm environments, the skin blood vessels dilate (vasodilation) which increases skin blood flow and results in increased skin temperature [1]. Several researchers have investigated the changes in skin blood flow at different locations of the body and its correlations with thermal comfort sensations [9, 14, 34–36]. In particular, temperatures from different locations in the face, arms, wrist, trunk, legs and hands are useful in predicting thermal sensations [1, 9, 14, 30, 36]. Many of the previous studies relied on different wearable devices to monitor changes in skin temperature from different locations. However, due to its intrusive nature, wearable devices might not be acceptable by many occupants.

Previous studies show that temperatures from different regions of the face such as forehead, nose, ears, cheeks can be utilized for predicting thermal sensations [1, 9, 14]. Because face is usually not covered by items of clothing and temperatures from different regions of the face can be used to predict thermal sensations, recent studies have utilized infrared thermography as an alternative to monitor skin temperatures non-intrusively. For example, Ghahramani et al. fitted non-contact infrared temperature sensors in an eyeglass frame to collect temperatures from forehead, nose, ear and cheek, and used a hidden Markov model to build models that achieved 83% accuracy in predicting uncomfortable conditions [14]. Ranjan et. al. captured thermal images of the face and hands twice per day from 30 participants using a high resolution thermal camera (FLIR A655sc), manually extracted temperatures from the thermal images, and showed that temperatures from the face and hands can be used to predict individual thermal sensations [29]. In another study, Burzo et. al. used FLIR A40, a high resolution thermal camera to acquire thermal videos, extracted several features from each frame without specifying different ROIs, and showed that significant correlations existed between the extracted features and thermal sensations [6]. Although the studies showed the usefulness

of using infrared thermography, manual segmentation of thermal images or locating different ROIs manually is cumbersome and not feasible for implementation in actual buildings. Automated methods for extracting skin temperatures from different ROIs can support the deployment of infrared thermography on a larger scale.

Several approaches have been developed to automatically extract useful information from thermal images. Pavlin et. al. utilized a low resolution thermal camera (FLIR Lepton) to extract temperatures from the forehead region and showed that the extracted temperatures were well correlated with thermal sensations [27]. The approach utilized a low-resolution thermal camera however, it was only able to extract crude measurements from the forehead. Li et. al. used the Haar-Cascade classifier using the Viola-Jones algorithm [33] to identify a bounding box surrounding the face in an RGB image, and identified different ROIs such as forehead, nose and cheeks based on relative location to the bounding box, and mapped the corresponding points to the thermal image to extract temperatures from different ROIs [23]. Using RGB data provides the opportunity to leverage advanced computer vision algorithms that fail to detect faces directly on the thermal images. In another study, Burzo et. al. also used the Haar-Cascades classifier using the Viola-Jones algorithm to detect face bounding boxes in the RGB image and mapped it to the thermal image to extract temperature from the entire face region [6]. The limitation of the approach by both Li et. al. [23] and Burzo et. al. [6] is that the Haar-Cascade classifier does not detect faces in an image when the head has some rotation or tilt relative to the camera. Furthermore, the location of different ROIs relative to the bounding box changes when the face orientation changes which can result in very noisy measurements. Although these previous approaches [6, 23] can detect occupant's faces and roughly locate the ROIs when the person is directly facing the camera in a frontal orientation, it is not practical to expect occupants to always be facing the camera in a frontal orientation. Some natural head tilt and rotation can lead to failure in detecting faces and inaccurate calculation of different ROI locations with previously studied approach of using Haar-Cascade classifiers. When occupants are outside the field of view of the cameras, no measurements can be made, and it is not practical to expect occupants to always be in the field of view of the cameras. However, having a better method to detect faces and locate different ROIs under some head tilt and orientation can help to improve thermal comfort predictions by extracting more and better temperature measurements from the thermal images when the occupants are in the camera's field of view.

To improve the detection of the face and locating different ROIs, Cosma et. al. used RGB-D (color and depth) data from a Kinect camera to recognize different body segments (head, torso, hand, shoulder etc.), and combined the information from FLIR Lepton thermal camera to extract skin and clothing temperatures from different segments [10]. The study showed that the variance in extracted skin and clothing temperatures could be used to predict thermal sensations. Metzmacher et. al. developed an approach to combine RGB-D data from a Kinect camera and thermal images from a high resolution FLIR A35 camera to extract skin temperatures from different facial ROIs [26]. The approach involved identifying a facemask in the Kinect data and overlaying it over the thermal image to extract skin temperatures. The facemask from the Kinect

data can be detected even when the person is not directly facing the camera and addresses the issue of an occupant having different orientations relative to the camera. In another study, Li et. al. developed an approach to use a network of multiple Kinect cameras and FLIR lepton cameras to capture images from different angles for the same occupant and extract overall temperature of the face [24]. However, adding an RGB-D camera increases the deployment costs. The low-resolution thermal camera, FLIR Lepton costs about \$200 and a Kinect camera costs over \$200. High resolution thermal cameras used in other studies cost around \$5000. A higher deployment cost can reduce the scalability of using thermal imaging for thermal comfort prediction.

In this study, we try to improve the extraction of skin temperatures from different ROIs in the face by combining data from a FLIR Lepton thermal camera and an inexpensive RGB camera. We leverage an algorithm for facial landmark detection, which can detect different locations in the face even when face is not oriented directly in front of the camera and provides a more robust way to calculate location of different ROIs on the face under different head tilt and rotation. The goal is to improve the quantity and quality of skin temperature extraction from different ROIs compared to previous studies by increasing the range of head tilt and rotation where different faces can be successfully detected, and improving the calculation of ROI locations in the face to reduce the noise in temperature measurements. We also compare our approach with ROIs detected using the Haar-Cascade classifier using the Viola-Jones algorithm used in previous studies.

3 ROI TEMPERATURE EXTRACTION

The overall process of extracting skin temperature from different ROIs involves image registration between the RGB and thermal images, facial landmark detection and calculation of ROI locations in the RGB image, mapping of ROI locations to the thermal image, and extracting temperatures from the thermal image. The images are collected using a regular RGB camera and a FLIR Lepton thermal camera shown in Figure 1. FLIR Lepton is a low-cost thermal camera which is capable of taking 80×60 pixel thermal images and has an accuracy of $\pm 5^\circ\text{C}$ and a resolution of 0.1°C for temperature measurements. Although higher resolution thermal cameras are available, FLIR Lepton was selected due to its low cost and thus its potential to be deployed at a larger scale. FLIR Lepton was previously validated by Li et. al. [23] against the FLIR T450SC camera, which is capable of capturing 320×240 pixel thermal images with an accuracy of $\pm 2^\circ\text{C}$. The authors showed that the FLIR Lepton camera was adequate for monitoring changes in skin temperature for thermal comfort assessment. The FLIR lepton camera is factory calibrated to provide accurate measurements and automatically performs recalibration when the sensor temperature changes [13], therefore, sensor calibration was not performed in this study.

3.1 Image Registration

As seen in Figure 1, there is a small gap between the RGB camera and the thermal camera. The RGB camera and the thermal camera capture the face from slightly different angles and resolution. The FLIR lepton is a low resolution thermal camera that captures 60×80 pixel images, and the RGB camera captures 600×800 pixel

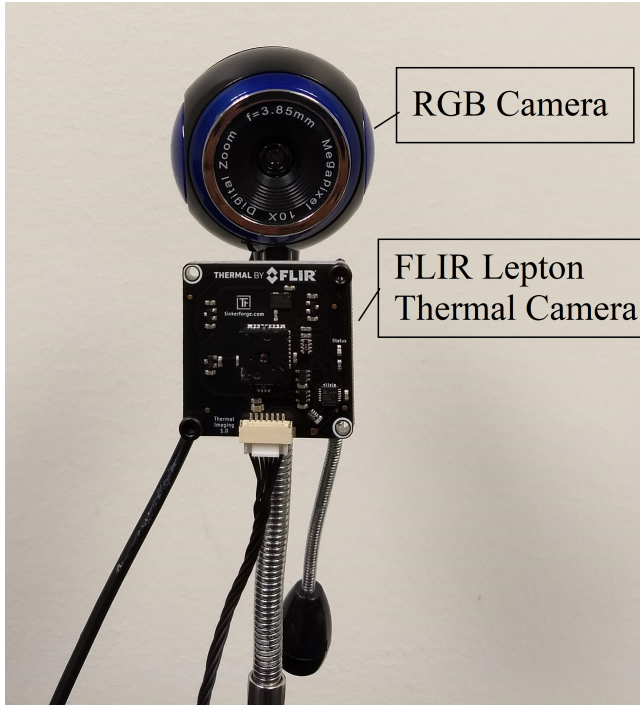


Figure 1: Camera setup showing the RGB and thermal cameras

images. To help the image registration process the thermal images are scaled up by a factor of 10 to 600×800 pixels using bicubic interpolation where the output pixel value is a weighted average of pixels in the nearest 4×4 neighborhood. It is important to note that the resized thermal images are only used for image registration and not for taking temperature measurements because the interpolation might lead to an incorrect scaling of temperature values. In order to map the RGB image with the thermal image, the RGB image is transformed to match the thermal image coordinates. Depending on the gap and orientation of the RGB and thermal cameras, the RGB image might need some scaling, rotation and translation to match the thermal image. In order to compute the relevant transformation matrix, some corresponding control points are manually defined for the setup. Figure 2 shows an example of the control points defined for our camera setup. The control points are defined manually by identifying the same point in two images. Although this is a manual process, it only needs to be performed once for the camera setup as long as the physical angles and distances between the two cameras do not change. A larger number of control points can reduce the human error resulting from the manual identification of the control points. The transformation matrix is computed using the following equation:

$$\begin{bmatrix} u & v \end{bmatrix} = \begin{bmatrix} x & y & 1 \end{bmatrix} \times \begin{bmatrix} sc & -ss \\ ss & sc \\ tx & ty \end{bmatrix}$$

$$sc = scale * \cos(angle)$$

$$ss = scale * \sin(angle)$$

Where u,v are the coordinates of the transformed point, x,y are the coordinates of original point before transformation. The scaling factors and rotation between the two images is captured by two variables sc and ss . The translation between the two images is captured by the variables tx and ty . The image registration process requires input of several (u,v) and (x,y) pairs of control points, and calculates sc , ss , tx and ty to find the transformation matrix. Although four control points are sufficient to calculate the four variables, it is generally recommended to provide more control points to account for human error during the control points selection shown in Figure 2.

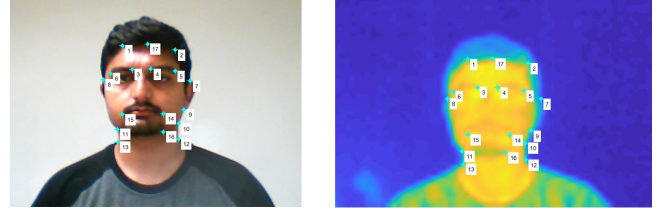


Figure 2: Manual selection of control points for image registration

3.2 ROI Location and Temperature Extraction

After the image registration, we run the facial landmark detection algorithm by Kazemi et. al. [20] using the dlib library. Dlib is an open source general purpose software library implemented in C++ and contains software components spanning various domains. Other off the shelf implementations of facial landmark detection are available in the openCV library, which is a library of components dedicated to computer vision. However, the dlib implementation was used in this study because it has a higher accuracy of facial landmark detection compared to the implementations available in the openCV library [17]. The algorithm detects 68 different points in the face shown in Figure 3. The desired ROIs can then be calculated based on their relative location to the 68 detected landmarks. Due to the high level of detail in the locations of 68 points in the face, this approach can provide a more robust identification of different ROIs compared to previous methods. Because of the level of detail achieved using facial landmarks, it has been used in different applications such as pose estimation, expression analysis, face identification [17].

Once the facial landmarks are identified, the location of ROIs in this study is calculated as follows. The nose center is calculated by averaging the location of points in the nose, points 28 to 31 in Figure 3. The x coordinate of the forehead center is calculated by averaging x coordinates of points 22 and 23 on the left and right eyebrows, and y coordinate of the forehead center is calculated by adding the vertical distance between the eyebrows points 22 and 23 and the nose center to the y coordinate of the eyebrows in the upward direction. The resulting location of the forehead center is slightly above the midpoint of the two eyebrows. The left cheek point is located by averaging points 15 and 34, which finds a point between the center of the nose and boundary of the left cheek. Similarly, the right cheek is located by averaging points 3 and 34 which finds a point between the center of the nose and boundary

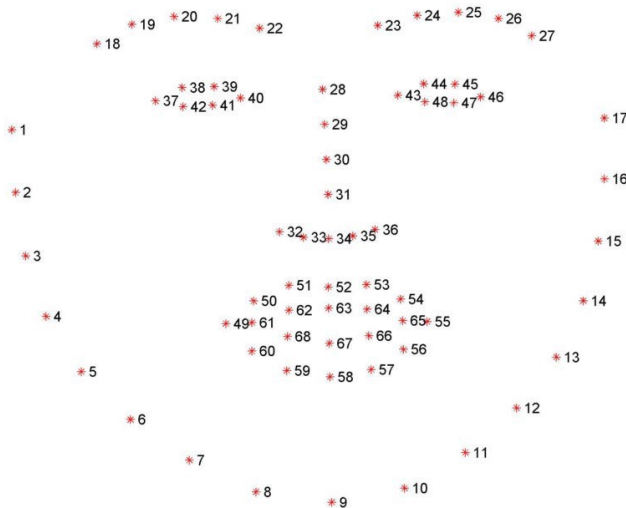


Figure 3: Points identified by the facial landmark detector [20]

of the right cheek. The facial landmark detection is performed on the RGB images that are in the same coordinates as the enlarged thermal image. The calculated location of ROIs is scaled down by a factor of 10 to match the original thermal image. The original thermal image of 60×80 pixels consists of direct temperature measurements, hence the temperatures of the located ROIs are extracted from the original thermal image.

4 EVALUATION

We directly compare our approach of using facial landmarks to locate different ROIs against using Haar-Cascade classifier used in previous studies to evaluate the quantity and quality of the extracted data. We first conducted a preliminary evaluation to investigate the range of facial orientations where different ROIs can be identified using the facial landmark detection algorithm and by the Haar-Cascade classifier. We then evaluate the two approaches on data collected from 20 participants in an experiment where the room temperature was gradually changed between 19°C and 29°C to explore the feasibility of extracting skin temperature from the ROIs under different thermal conditions while performing activities on a computer. The ROIs used in this study are forehead, nose, left cheek and right cheek because they were identified as useful sites to monitor skin temperature in previous studies [14, 23].

4.1 Preliminary Evaluation

In order to explore the range of face orientations where facial landmark detection and the Haar-Cascade classifier are able to detect faces, we first collected images from the RGB camera and the thermal camera for about 6 minutes. A frame was simultaneously captured from both cameras every second resulting in a total of 363 sets of images. In this preliminary evaluation, the participant gradually moved his head from left to right, up to down, and along two diagonal directions relative to the camera. We then ran the pre trained facial landmark detection [20] from the dlib library, and a

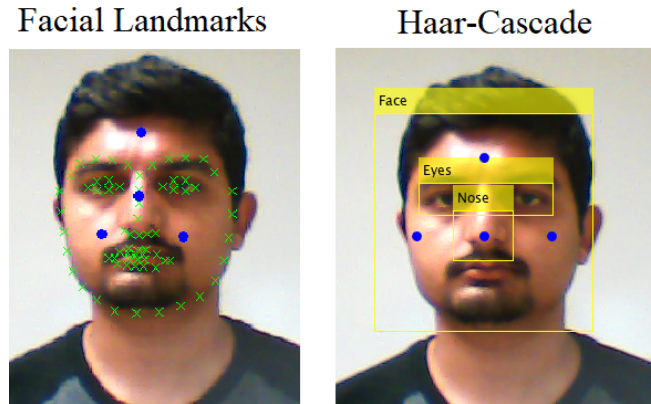


Figure 4: Illustration of facial landmarks (left: green crosses), bounding boxes from the Haar-Cascade classifier (right: yellow boxes) and calculated ROI points (both images: blue circles)

pre-trained Haar-Cascade classifier using the Viola-Jones algorithm [33] on the collected RGB images. The facial landmark detection algorithm detects boundary points along the face, as well as eyes, nose and mouth. The calculation of ROIs from facial landmarks is explained in section 3.2. The Haar-Cascade classifier only identifies a bounding box surrounding a particular region. Because the goal is to achieve accurate identification of different ROIs, we run the Haar-Cascade classifier three times to detect the face, eyes and nose. We then calculate the ROIs based on their relative locations to the detected points. For the Haar-Cascade classifier, the nose is calculated as the center of the bounding box surrounding the nose. The two cheeks are calculated by taking half of the width of the bounding box surrounding the eyes on either side of the nose center point. The forehead is calculated by taking a point aligned to the nose center that is above the eyes bounding box by half of the width of the nose bounding box. Figure 4 shows the detected facial landmarks, bounding boxes and calculated ROI locations for the two algorithms.

As seen in Figure 4, both approaches can be used to locate different ROIs when the face is oriented directly in front of the camera. We then evaluated both approaches under different face orientations. Facial landmarks detection algorithm successfully detected the landmarks in 313 out of 363 images, whereas the Haar-Cascade classifier only detected faces in 159 out of 363 images. Figure 5 provides a visual comparison of the two algorithms and their capabilities to detect faces under different orientations. The central images in Figure 5 (shown inside a black boundary) show the blended images of the frames where the algorithms successfully detected the face in the image. The eight images on the sides of the center images in Figure 5 show the blended images of frames where the algorithms failed to detect faces in each side orientation. A lighter image indicates lower number of frames and a darker image indicates a larger number of frames. A white image indicates that faces were detected for all of the frames in that orientation. As seen from the eight side images, the Haar-Cascade algorithm fails to

Table 1: Anthropometric details of study participants

Gender	Count	Age(years)	Height(cm)	Weight(kg)
Male	12	23.9 \pm 4.6	178.6 \pm 6.8	80.0 \pm 13.0
Female	8	29.5 \pm 14.8	161.6 \pm 7.2	64.5 \pm 16.5
Overall	20	26.2 \pm 10.1	171.8 \pm 10.9	73.8 \pm 16.1

detect faces in a lot more frames compared to the facial landmark detection algorithm.

4.2 Data Collection

From our preliminary evaluation, we see that Facial Landmarks algorithm can successfully detect faces in more orientations compared to the Haar-Cascade classifier. For a more comprehensive evaluation, we collected data from 20 participants in a realistic office setting while they were using their computers. The data were collected in a research office at the University of Southern California (USC) in Los Angeles during the summer months of June 2018 to August 2018. All of the participants were asked to wear pants and t-shirts during the experiment to keep the clothing levels consistent. The experiment procedure was explained to the participants and informed consent was obtained before starting the experiment. The study was approved by the Institutional Review Board (IRB) at USC. Twenty healthy subjects, 12 males and 8 females, participated in the study. Table 1 shows the anthropometric details of the study participants.

The experiment was conducted in two separate segments: cold segment and hot segment. Each segment lasted between 1 to 1.5 hours. In the cold segment the temperature was gradually decreased from roughly 24°C to 19°C, and in the hot segment the temperature was gradually increased from roughly 22°C to 29°C. There was at least 1-hour gap between the segments where the room was adjusted to the new starting point and participants were transferred to another location with a neutral temperature during this period for the participants to re-acclimate to the neutral environment. The room temperature was gradually changed at the rate of roughly 1°C/10 mins to avoid sudden changes in the thermal environment during both segments. The room temperature was monitored every second using a DHT22 sensor connected to an Arduino Uno placed on the desk roughly 0.5m from the participant. The minimum air temperature and its standard deviation for the participants during the experiment was $19.3 \pm 1^\circ\text{C}$. The maximum air temperature and its standard deviation was $28.7 \pm 0.6^\circ\text{C}$. A regular webcam was used to capture RGB images of the participants' faces and a FLIR Lepton thermal camera was used to capture thermal images of the participant's faces every second. The camera setup is shown in Figure 1.

The participants were working on their computer during the study, and they were not asked to stay upright facing the camera. The dataset provides a realistic reflection of a person's movement while working on their computer compared to the previous studies where participants were asked to maintain specific poses [23, 29]. In order to understand how the two algorithms may perform in real life, we use the dataset collected from 20 participants during our experiment to evaluate the percentage of images where each

algorithm successfully detects different regions of the face to indicate the quantity of useful information extracted by each algorithm. We then evaluate the noise in skin temperature measurements extracted from the two algorithms to indicate the quality of useful information obtained from the two algorithms.

4.3 Results

Overall, 7700 sets of RGB and thermal images were collected on average from each participant, resulting in over 154,000 sets of images. For each participant, we evaluated the percentage of images where different sites in the face were successfully detected. The results for each participant is shown in Figure 7. The facial landmark algorithm was able to detect faces in 85% of images on average. The Haar-Cascade classifier was able to detect faces in 60% of images on average. In Figure 7, we see that both algorithms have low detection rates for some participants. This is caused by the downward head tilt of the participants because they were hunching while using their computers. The downward head tilt is a weak point for both algorithms as seen in our preliminary investigation shown in Figure 5.

Figure 7 shows the quantity of images successfully processed by the algorithms. The quality of the extracted temperatures is also important to consider. It is difficult to quantify the quality of temperature measurements without having a reference sensor attached to participants' faces. However, having a reference sensor attached to participant's faces would interfere with the temperature extraction from the thermal images. In this study, we assume that a signal smoothed using a moving average filter with a large window can be used as a reference signal. A moving average filter of size M calculates an output point by taking average of M neighboring points for every data point. In our case, the moving average filter removes the noise that is caused by small movements in participants' faces that can cause errors in the calculation of different ROIs. We then calculate the residual error by subtracting the raw temperature measurements from the moving average.

An example of raw temperature measurements and the smoothed temperature measurements using moving average filter of size 200 from the forehead of a participant is shown in Figure 6. We then compute the average and standard deviation of the residual error in all four ROIs for each participant to get a general idea of the level of noise in the temperature measurements. The distribution of residual errors is shown in Figure 8 where the whiskers of the box plots indicate average ± 1.5 standard deviation. The outliers are not shown in Figure 8 to improve the clarity of the figure. It is important to note that because we are using a moving average smoothed signal as reference, the residual average errors are close to 0. The standard deviation as seen from the length of the whiskers in the boxplots gives an indication of the fluctuation in temperature measurements resulting from the errors in the ROI calculation. In Figure 8 we see that the noise in temperature measurements extracted using the Haar-Cascade classifier is higher than in temperature measurements extracted using the facial landmarks detection. In general, the noise is higher for the participants who moved more during the data collection and who had a hunched posture. The noise in temperature measurements for the four ROIs is quite similar when the ROIs are calculated using facial landmark. This is because the

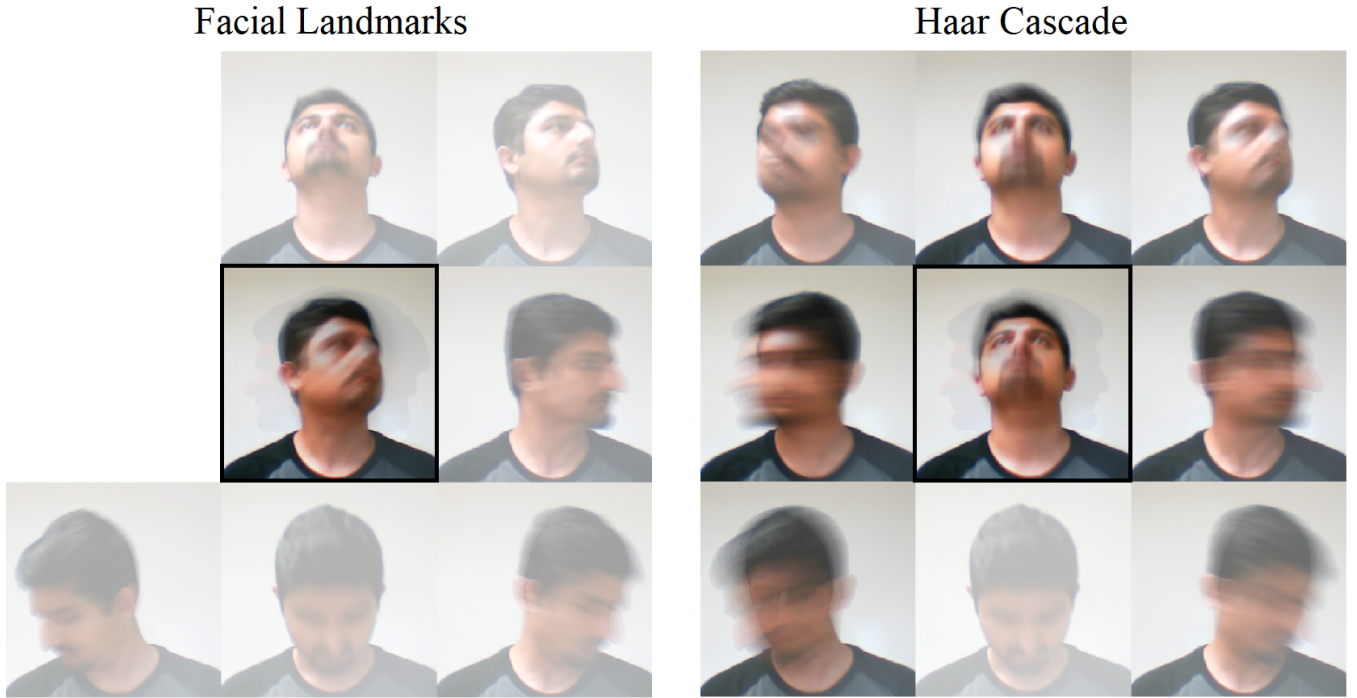


Figure 5: Blended images of frames with failed detection in each orientation (8 sides) and successful detection (center) for each algorithm

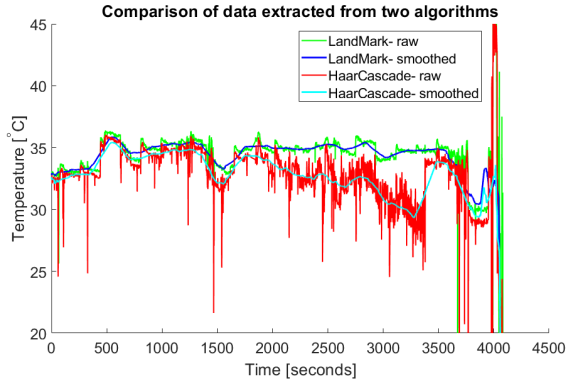


Figure 6: Comparison of forehead temperatures extracted from two algorithms

facial landmark can provide a more granular location of different points in the face. The noise in temperature measurements in the four ROIs calculated using the Haar-Cascade bounding boxes is quite different between the left cheek and the right cheek because the bounding boxes do not provide a good estimation of the face orientations. On average, the standard deviation in the residual error was about 1°C for the facial landmarks approach and about 2°C for the Haar-Cascade approach. This indicates that our approach can extract more accurate measurements from the thermal images compared to previous methods.

5 DISCUSSION

Thermal imaging has recently gained interest as a non-intrusive method to monitor physiological data for thermal comfort assessment. Several studies have shown the usefulness of thermal imaging for predicting thermal sensations. In this study, we focused on a new method to leverage facial landmark detection to accurately calculate different ROIs in the face and extract corresponding skin temperatures from thermal images. Our results indicate that facial landmark detection can more accurately locate different ROIs in the face compared to the Haar-Cascade classifier used in previous studies. If such system is to be used in real buildings, it is not reasonable to expect occupants to always be facing the camera. Our approach enables us to extract more information from the frames where occupants are facing the camera. Although better than the previous approaches, our approach still has some limitations. Facial landmark detection works well when the faces are directly facing the camera and head tilt, or head rotation is relatively small. However, Better ways to locate different ROIs in non-frontal faces can help to improve the extraction of skin temperatures from thermal images. Furthermore, ways to handle missing measurements should be incorporated into a system that is designed for practical implementation. In our study, we noticed that some of the participants were naturally hunched while using their computer, which resulted in a high number of images with downward head tilt where facial landmarks were not detected. For occupants who have a naturally hunched posture, a better placement of the camera that accounts for the head tilt can help address the issue.

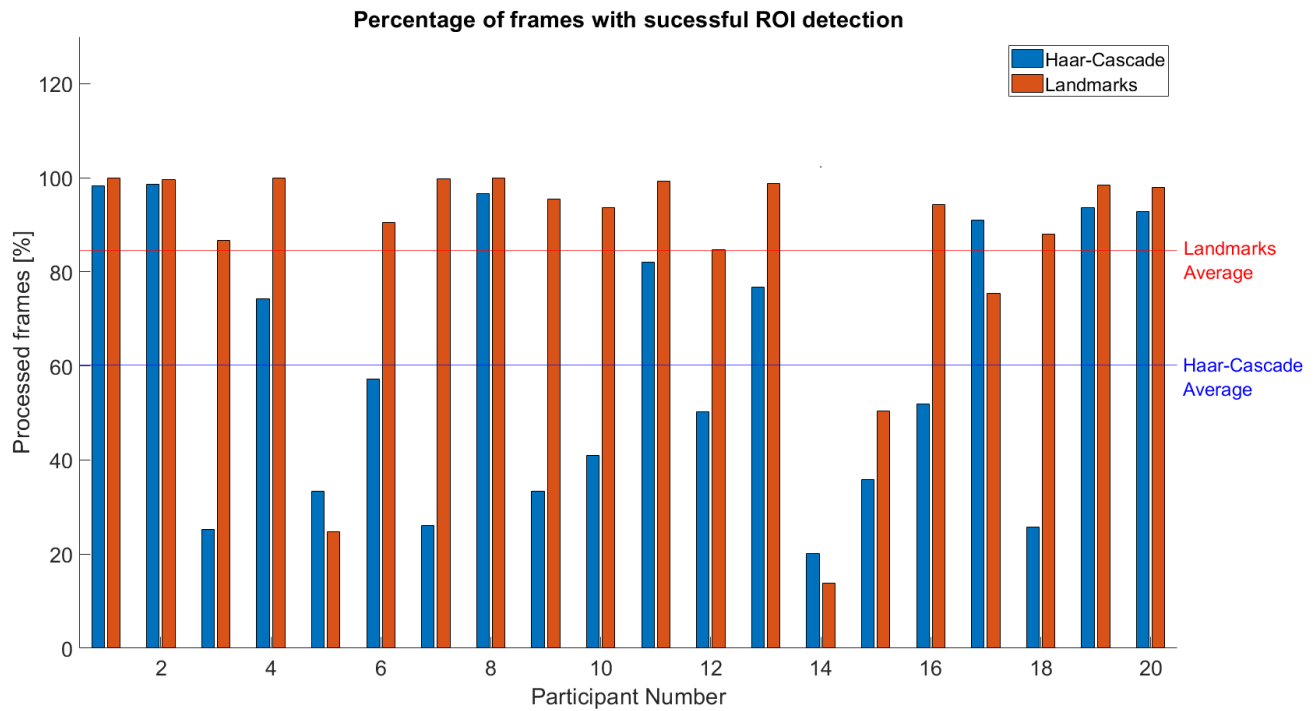


Figure 7: Percentage of frames with successful ROI detection for each participant

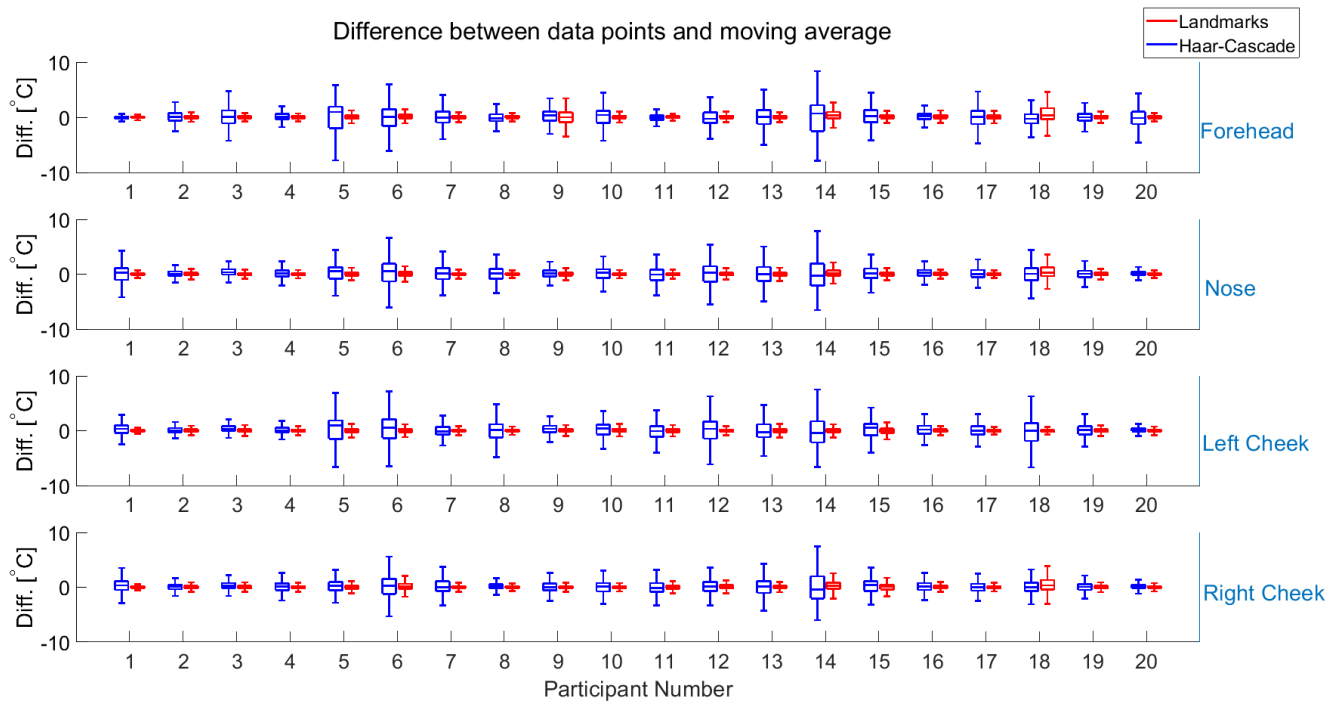


Figure 8: Difference between each data point and moving average (residual error) for each participant

One of the challenges during this study was to quantify the noise or errors in the temperature measurements. In this study, we used a moving average signal with a large window as the reference measurement. A better way to quantify the errors can provide a better understanding of the quality of measurements. For example, some markers could be placed near the desired ROIs and temperature near the markers could be extracted for validation. However, because the comparison of both algorithms is performed on the same exact frames, and the dataset is relatively large consisting people of different age, gender and body compositions, we can expect facial landmark detection to provide data with lower noise in a real world implementation as well.

In this study, we utilized a low cost thermal camera and a RGB camera. Because both the facial landmark detection and Haar-Cascade algorithms utilize RGB images, the results related to detection and location of ROIs in the face is independent of the thermal camera used. Using a higher accuracy thermal camera can lead to more accurate temperature measurements but will increase the deployment costs. Some of the previous studies also leveraged RGB-D data from a Kinect camera, which also provides depth information in addition to the RGB information. The inclusion of depth information can help in a more robust identification of different locations on face. However, including depth information requires additional deployment costs. The Kinect camera, which is the most widely used depth camera costs around \$200. The FLIR Lepton, a low-cost thermal camera used in this study also costs around \$200. Including a depth camera could double the deployment costs of the system. Furthermore, a future study to directly compare the accuracy trade-offs between using RGB-D data vs RGB data for calculating ROIs is needed to understand the accuracy tradeoffs between the two approaches.

The RGB and thermal cameras were registered in this study by manually defining control points between the two images. Although this registration only needs to be performed once for each camera setup, it is subject to human errors. Defining a larger number of control points manually can reduce human errors. Several automatic camera registration techniques also exist in literature and can be used to reduce errors in the camera registration process [24, 26]. The automatic camera registration process involves using checkerboards, circular grids, or other repeated patterns where the boundaries can be automatically detected and used as control points. The additional requirements are: the checkerboard patterns need to have different thermal emissivity and temperatures for the thermal image to clearly capture the control points.

This study focused on the evaluation of the actual temperature measurements because previous studies have shown the usefulness of skin temperature measurements from different ROIs on face for predicting thermal sensations with 75%-85% accuracy [3, 6, 23]. For the dataset used in this study, the accuracy of predicting thermal sensations from facial skin temperatures extracted using facial landmark detection resulted in an accuracy of 76%. The accuracy improved to 85% when ambient air temperature was used in addition to facial skin temperature. Readers are referred to another paper by the authors [3] which focuses on predicting thermal sensations using the same dataset as this paper. One of the limitations of this study is that we did not compare the impact of the noise in temperature measurements on predicting thermal sensations.

Future studies are necessary to quantify the quality of temperature measurements automatically extracted from the thermal images and the impact of noise in the measurements on modeling of thermal comfort sensations. Furthermore, the actual performance of personal comfort models when used to control the thermal environment, and its influence on occupant comfort has not been explored well in the literature. Extensive studies are required to evaluate the performance of personalized comfort models in controlling the thermal environment and improving occupant satisfaction.

From a practical perspective, although thermal imaging is non-intrusive, there are still privacy concerns surrounding the use of cameras. Clear disclosure of what data is collected and how it is used might help alleviate some of the privacy concerns. Furthermore, processing the camera data in real time to extract relevant information and not storing the images or video feed might help to alleviate some of the privacy concerns.

6 CONCLUSION

Infrared thermography or thermal imaging provides a non-intrusive way to monitor changes in skin temperature for thermal comfort assessment. Previous studies have shown the usefulness of extracting skin temperatures from different ROIs in the face to predict thermal sensations. One of the major challenges of using thermal imaging for monitoring skin temperatures in actual buildings is that occupants may move relative to the camera. It is not practical to expect building occupants to be oriented facing the cameras at all times, therefore, it is important to be able to extract as much information as possible from instances where it is feasible to extract relevant information. In this paper we described a novel approach to locate different ROIs in the face by leveraging recent advancements in computer vision in RGB images, and to extract temperatures of those ROIs from thermal images. We compared our approach of using facial landmark detection to locate different ROIs against detecting bounding boxes around different sections of the face with Haar-Cascade classifiers used in previous studies. In this study, we used a dataset of over 154,000 RGB and thermal images and evaluated our approach against previous approach used in the literature. The facial landmark algorithm was successful in identifying 85% of the faces in our dataset compared to 60% successful detection with the Haar-Cascade classifiers. We also evaluated the level of noise in the temperature measurements from the two approaches. Our approach was able to reduce the level of noise by a factor of 2 compared to previous approach. Our results show that facial landmark detection can provide more accurate temperature extraction from different ROIs in the thermal image compared to the Haar-Cascade classifiers. Our approach has the potential to improve the accuracy of thermal comfort prediction using physiological measurements. One of the major limitations of this study is that the facial landmark detection algorithm does not perform well when the head is tilted or rotated far away from the frontal position. Future studies are needed to improve the calculation of ROI locations under different face orientations.

ACKNOWLEDGMENTS

This material is based upon the work supported by the National Science Foundation under Grant No. 1763134 and 1351701. Any

opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The help of research assistants Paulina Maldonado, Eddy Solares Quezada, Irie Cooper and Victoria Sanchez during sensor setup and data collection is greatly appreciated.

REFERENCES

- [1] E. Arens and H. Zhang. 2006. The skin's role in human thermoregulation and comfort. In *Thermal and Moisture Transport in Fibrous Materials*. Woodhead Publishing Ltd, Berkeley, CA, 560–602. <https://doi.org/10.1533/9781845692261.3.560>
- [2] Ashrant Aryal and Burcin Becerik-Gerber. 2018. Energy consequences of Comfort-driven temperature setpoints in office buildings. *Energy and Buildings* 177 (aug 2018), 33–46. <https://doi.org/10.1016/j.enbuild.2018.08.013>
- [3] Ashrant Aryal and Burcin Becerik-Gerber. 2019. A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor. *Building and Environment* 160 (aug 2019), 106223. <https://doi.org/10.1016/j.buildenv.2019.106223>
- [4] Ashrant Aryal, Burcin Becerik-Gerber, Francesco Anselmo, Shawn C. Roll, and Gale M. Lucas. 2019. Smart Desks to Promote Comfort, Health, and Productivity in Offices: A Vision for Future Workplaces. *Frontiers in Built Environment* 5 (jun 2019), 76. <https://doi.org/10.3389/fbuil.2019.00076>
- [5] ASHRAE. 2017. *ASHRAE Standard 55-2017: Thermal Environmental Conditions for Human Occupancy*. Technical Report. ASHRAE. <https://www.ashrae.org/technical-resources/standards-and-guidelines/read-only-versions-of-ashrae-standards>
- [6] Mihai Burzo, Cakra Wicaksono, Mohamed Abouelenien, Veronica Perez-Rosas, Rada Mihalcea, and Yong Tao. 2014. Multimodal sensing of thermal discomfort for adaptive energy saving in buildings. In *iSBE NET ZERO BUILT ENVIRONMENT*. 344.
- [7] Tanaya Chaudhuri, Deqing Zhai, Yeng Chai Soh, Hua Li, and Lihua Xie. 2018. Thermal comfort prediction using normalized skin temperature in a uniform built environment. *Energy and Buildings* 159 (jan 2018), 426–440. <https://doi.org/10.1016/j.enbuild.2017.10.098>
- [8] Toby Cheung, Stefano Schiavon, Thomas Parkinson, Peixian Li, and Gail Brager. 2019. Analysis of the accuracy on PMV and PPD model using the ASHRAE Global Thermal Comfort Database II. *Building and Environment* (feb 2019). <https://doi.org/10.1016/j.buildenv.2019.01.055>
- [9] Joon-Ho Choi and Dongwoo Yeom. 2017. Study of data-driven thermal sensation prediction model as a function of local body skin temperatures in a built environment. *Building and Environment* 121 (2017), 130–147. <https://doi.org/10.1016/j.buildenv.2017.05.004>
- [10] Andrei Claudiu Cosma and Rahul Simha. 2018. Thermal comfort modeling in transient conditions using real-time local body temperature extraction with a thermographic camera. *Building and Environment* 143 (oct 2018), 36–47. <https://doi.org/10.1016/j.buildenv.2018.06.052>
- [11] David Daum, Frédéric Haldi, and Nicolas Morel. 2011. A personalized measure of thermal comfort for building controls. *Building and Environment* 46, 1 (jan 2011), 3–11. <https://doi.org/10.1016/j.buildenv.2010.06.011>
- [12] R de Dear, GS Brager, and Donna Cooper. 1998. *Developing an adaptive model of thermal comfort and preference*. Technical Report Part 1. 1–18 pages. https://escholarship.org/uc/item/4qq2p9c6.pdf?%5Cnhttp://escholarship.org/uc/item/4qq2p9c6.pdf?%5Cnhttp://repositories.cdlib.org/cedr/cbe/ieq/deDear1998_1_ThermComPref
- [13] FLIR. 2016. *FLIR LEPTON® with Radiometry Datasheet*. Technical Report. www.flir.com/globalassets/imported-assets/document/lepton-engineering-datasheet---with-radiometry.pdf
- [14] Ali Ghahramani, Guillermo Castro, Simin Ahmadi Karvigh, and Burcin Becerik-Gerber. 2018. Towards unsupervised learning of thermal comfort using infrared thermography. *Applied Energy* 211 (feb 2018), 41–49. <https://doi.org/10.1016/j.apenergy.2017.11.021>
- [15] Switzerland International Standard Organization, Geneva. 2005. *ISO 7730:2005(en), Ergonomics of the thermal environment – Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria*. Technical Report. ISO. <https://www.iso.org/obp/ui/#iso:std:iso:7730:ed-3:v1:en>
- [16] Farrokh Jazizadeh, Ali Ghahramani, Burcin Becerik-Gerber, Tatiana Kichkaylo, and Michael Orosz. 2014. Human-Building Interaction Framework for Personalized Thermal Comfort-Driven Systems in Office Buildings. *Journal of Computing in Civil Engineering* 28, 1 (jan 2014), 2–16. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000300](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000300) arXiv:arXiv:1011.1669v3
- [17] Benjamin Johnston and Philip de Chazal. 2018. A review of image-based automatic facial landmark identification techniques. *EURASIP Journal on Image and Video Processing* 2018, 1 (dec 2018), 86. <https://doi.org/10.1186/s13640-018-0324-4>
- [18] Wooyoung Jung and Farrokh Jazizadeh. 2018. Vision-based thermal comfort quantification for HVAC control. *Building and Environment* 142 (sep 2018), 513–523. <https://doi.org/10.1016/j.buildenv.2018.05.018>
- [19] Caroline Karmann, Stefano Schiavon, and Edward Arens. 2018. Percentage of commercial buildings showing at least 80% occupant satisfied with their thermal comfort. *Proceedings of 10th Windsor Conference: Rethinking Comfort*. <https://cloudfront.escholarship.org/dist/prd/content/qt89m0z34x/qt89m0z34x.pdf?t=p9m07w>
- [20] Vahid Kazemi and Josephine Sullivan. 2014. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1867–1874. <https://doi.org/10.1109/CVPR.2014.241>
- [21] Joyce Kim, Stefano Schiavon, and Gail Brager. 2018. Personal comfort models – A new paradigm in thermal comfort for occupant-centric environmental control. *Building and Environment* 132 (jan 2018), 114–124. <https://doi.org/10.1016/j.buildenv.2018.01.023>
- [22] Changzhi Li, Victor M. Lubecke, Olga Boric-Lubecke, and Jenshan Lin. 2013. A Review on Recent Advances in Doppler Radar Sensors for Noncontact Healthcare Monitoring. *IEEE Transactions on Microwave Theory and Techniques* 61, 5 (may 2013), 2046–2060. <https://doi.org/10.1109/TMTT.2013.2256924>
- [23] Da Li, Carol C. Menassa, and Vineet R. Kamat. 2018. Non-Intrusive Interpretation of Human Thermal Comfort through Analysis of Facial Infrared Thermography. *Energy and Buildings* (jul 2018). <https://doi.org/10.1016/j.enbuild.2018.07.025>
- [24] Da Li, Carol C. Menassa, and Vineet R. Kamat. 2019. Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras. *Applied Energy* 251 (oct 2019), 113336. <https://doi.org/10.1016/j.apenergy.2019.113336>
- [25] Shichao Liu, Ming Jin, Hari Prasanna Das, Costas J Spanos, and Stefano Schiavon. 2018. Personal thermal comfort models based on physiological parameters measured by wearable sensors. In *Proceedings of the Windsor Conference*. 431–441.
- [26] Henning Metzmacher, Daniel Wölki, Carolin Schmidt, Jérôme Frisch, and Christoph van Treeck. 2018. Real-time human skin temperature analysis using thermal image recognition for thermal comfort assessment. *Energy and Buildings* 158 (jan 2018), 1063–1078. <https://doi.org/10.1016/j.enbuild.2017.09.032>
- [27] Boris Pavlin, Giovanni Pernigotto, Francesca Cappelletti, Paolo Bison, Renato Vidoni, and Andrea Gasparella. 2017. Real-Time Monitoring of Occupants' Thermal Comfort through Infrared Imaging: A Preliminary Study. *Buildings* 7, 4 (feb 2017), 10. <https://doi.org/10.3390/buildings7010010>
- [28] P.O. Fanger. 1970. *Analysis and Applications in Environmental Engineering*. Danish Technical Press (1970), 244. <https://www.cabdirect.org/cabdirect/abstract/19722700268>
- [29] Juhi Ranjan and James Scott. 2016. ThermalSense: Determining Dynamic Thermal Comfort Preferences using Thermographic Imaging. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '16*. ACM Press, New York, New York, USA, 1212–1222. <https://doi.org/10.1145/2971648.2971659>
- [30] Tomonori Sakoi, Kazuyo Tsuzuki, Shinsuke Kato, Ryozo Ooka, Doosam Song, and Shengwei Zhu. 2007. Thermal comfort, skin temperature distribution, and sensible heat loss distribution in the sitting posture in various asymmetric radiant fields. *Building and Environment* 42, 12 (dec 2007), 3984–3999. <https://doi.org/10.1016/j.buildenv.2006.10.050>
- [31] Soo Young Sim, Myung Jun Koh, Kwang Min Joo, Seungwoo Noh, Sangyun Park, Youn Ho Kim, and Kwang Suk Park. 2016. Estimation of thermal sensation based on wrist skin temperatures. *Sensors (Switzerland)* 16, 4 (mar 2016), 420. <https://doi.org/10.3390/s16040420>
- [32] US EIA. 2012. *Commercial Buildings Energy Consumption Survey*. *US Energy Information Administration* (2012).
- [33] P. Viola and M. Jones. 2005. 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. IEEE Comput. Soc, I–511–I–518. <https://doi.org/10.1109/cvpr.2001.990517>
- [34] Jing Xiong, Xin Zhou, Zhiwei Lian, Jianxiong You, and Yanbing Lin. 2016. Thermal perception and skin temperature in different transient thermal environments in summer. *Energy and Buildings* 128 (sep 2016), 155–163. <https://doi.org/10.1016/j.enbuild.2016.06.085>
- [35] H. Zhang. 2008. *Human thermal sensation and comfort in transient and non-uniform thermal environments*. Ph.D. Dissertation. <https://escholarship.org/uc/item/11m0n1wt>
- [36] Hui Zhang, Edward Arens, Charlie Huizenga, and Taeyoung Han. 2010. Thermal sensation and comfort models for non-uniform and transient environments, part III: Whole-body sensation and comfort. *Building and Environment* 45, 2 (feb 2010), 399–410. <https://doi.org/10.1016/j.buildenv.2009.06.020>