A Low-Cost and Scalable Personalized Thermal Comfort Estimation System in Indoor Environments

Peter Wei
pw2428@columbia.edu
Columbia University

Yanchen Liu
yl4189@columbia.edu
Columbia University

Hengjiu Kang
hk3120@columbia.edu
Columbia University

Chenye Yang
cy2540@columbia.edu
Columbia University

Xiaofan Jiang
jiang@ee.columbia.edu
Columbia University

ABSTRACT

In commercial buildings, occupant thermal comfort is a key factor that must be optimized to provide a comfortable and productive work environment. However, current methods largely estimate thermal comfort based on preset models which do not incorporate real-time measurements or individual thermal preferences. In this work, we present a scalable system for estimating personalized thermal comfort using low-cost thermal camera based sensor nodes. This system extracts non-intrusive thermal measurements, is robust to different perspectives and environments, is easily deployable and low-cost, and can incorporate individual thermal feedback for more personalized thermal comfort estimates. In comparison with baseline methods, our system is able to improve thermal comfort estimates on the ASHRAE 7-point thermal sensation scale by 64% over baseline methods.

CCS CONCEPTS

• **Computer systems organization** → *Embedded and cyber-physical systems*; • **Hardware** → **Sensor applications and deployments**; • **Human-centered computing** → *User models.*

KEYWORDS

Comfort Estimation, Object Detection, Orientation Estimation

ACM Reference Format:
Peter Wei, Yanchen Liu, Hengjiu Kang, Chenye Yang, and Xiaofan Jiang. 2021. A Low-Cost and Scalable Personalized Thermal Comfort Estimation System in Indoor Environments. In *The First ACM International Workshop on Cyber-Physical-Human System Design and Implementation (CPHS '21), May 18, 2021, Nashville, TN, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3458648.3460006

1 INTRODUCTION

In commercial buildings, research has largely focused on energy optimization [4, 25, 26]; however, occupant thermal comfort is also a critical factor that impacts both health and productivity [9]. For example, studies such as [1, 22] have shown that changes in perceived thermal comfort can indicate changes in mental awareness and arousal, which are key factors in productivity. However, [10] found that only 2% surveyed of buildings meet the 80% occupant thermal satisfaction requirement defined by ASHRAE Standard 55 [21], suggesting that for most buildings, both measurement and optimization of thermal comfort can still be improved.

Standard methods for measuring occupant thermal comfort, such as Fanger's predicted mean vote model [7], or more recent adaptive models [6, 14], rely on estimates for different environmental factors to estimate average occupant thermal comfort. These environmental factor estimates, such as metabolic rate, clothing insulation, and air temperature, are estimated beforehand and not representative of actual conditions. Furthermore, these methods provide an *average* thermal comfort, not accounting for individual thermal preferences.

Studies such as [15, 23] have shown that various personal factors, such as clothing, metabolic level, and thermal preferences can cause differences in perceived thermal comfort. Recent studies have attempted to incorporate real-time environmental measurements and personalized measurements to build better models for personalized thermal comfort estimation. As an example, some studies [13, 17] have used thermal cameras to measure facial temperature features to estimate individual thermal comfort; however, these studies have left unaddressed challenges such as scalability to larger deployments, personalized thermal comfort models, and robustness against occupant behavior.

In this work, we develop a novel thermal comfort estimation system that is easily scalable, learns general as well as personal comfort models, and is robust to different facial perspectives. This system can be easily deployed in indoor spaces such as commercial buildings and used in a variety of thermal comfort related applications, such as HVAC control or comfort-based recommendations. As an example, personalized comfort estimates can allow systems to group occupants with similar thermal preferences to reduce overall thermal discomfort. Our contributions in this work are:

(1) We present the architecture, design, and implementation of a real-time comfort estimation system that extracts and integrates facial temperature features with environmental sensing to provide personalized comfort estimates.
(2) We present a novel facial temperature feature pipeline which is robust to changes in facial perspectives.
(3) We evaluate our system using a two-week thermal comfort study in a commercial building deployment and demonstrate accuracy improvements of up to 64% over baseline methods.

CPHS '21, May 18, 2021, Nashville, TN, USA

ACM ISBN 978-1-4503-8440-7/21/05…$15.00
https://doi.org/10.1145/3458648.3460006

2 RELATED WORKS

For many years, thermal comfort has been estimated by the widely accepted models such as the predicted mean vote (PMV) [7], and adaptive models [6, 14], and have been integrated into the ASHRAE Standard [21]. However, one major drawback of these models is that these models estimate the average thermal comfort, and does not differentiate for individual occupants. Different studies have shown that individuals have different thermal preferences [19, 24], and thus an aggregate model may achieve an average thermal comfort without satisfying individual thermal comfort preferences.

Recently, there have been a few works utilizing RGB-thermal cameras for comfort sensing [3, 8, 12, 13, 17]. However, each of these works have drawbacks which limit adoption in real buildings. In [13], the authors require the thermal cameras to be placed in front of the occupants at less than 2 meters away, thus restricting scalability to sense multiple occupants' thermal comfort. The authors also only utilize average face temperature, regardless of the perspective of the face; this can result in thermal measurement differences depending on which parts of the face are in view of the thermal camera. In [12, 17], the thermal cameras are required to be even closer to the occupants, and require the occupants to modify their behavior to allow for specific thermal measurement readings (such as directly facing the camera). In [3], the authors focus on accurately extracting facial thermal features, and use an off the shelf landmark detection algorithm from [11] which relies on a frontal view of the face for detection.

Alternatively, there are a number of studies which have achieved high comfort sensing accuracy using contact sensors. [5] used skin temperature sensors with environmental sensors to train a model for comfort estimation; [20] relies on a wrist temperature sensor to predict thermal sensation; [8] uses infrared sensors mounted on wearable glasses to measure facial skin temperatures. In these studies, sensors need to be contacted with each occupant, which is a major disadvantage for adoption in real deployments.

3 SYSTEM DESIGN

As shown in Figure 1, the system consists of a number of sensor nodes collecting thermal and RGB images at various locations, as well as environmental sensors to provide additional features such as temperature and humidity. These data streams constitute the main features for estimating personal thermal comfort. We include a web interface for occupants to provide perceived thermal comfort feedback, which is used as the target for training comfort models. The features and feedback are processed at a central server, which is responsible for learning general and personalized comfort models.

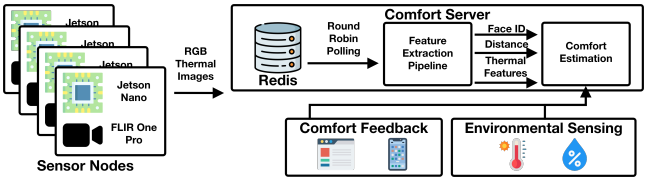

Figure 1: System Architecture, consisting of sensor nodes, comfort feedback interface, environmental sensing, and a central comfort server.

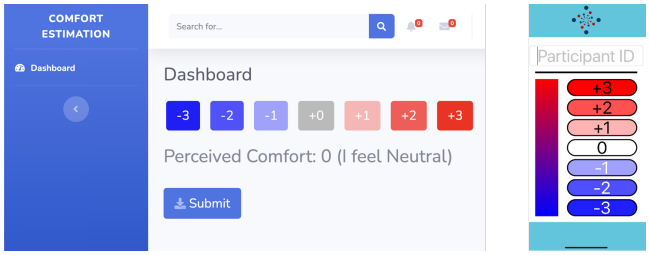

Figure 2: Web and mobile interface for submitting thermal comfort feedback.

3.1 Thermal Comfort Sensor Node

There are two criteria which are emphasized to improve deployability. First, each sensor node should be low-cost, to allow for the deployment of multiple sensors throughout a space. Second, each sensor node should require minimal configuration to allow for quick setup and easy addition of additional sensors nodes.

To minimize cost of deployment, the sensor node is composed of two main components: a FLIR One Pro thermal camera, and an NVIDIA Jetson Nano board. The FLIR One Pro thermal camera is significantly lower cost than most commercial thermal cameras, which often cost upwards of 1,000$. In addition, the Jetson Nano board replaces the typical mobile device interface for the FLIR One Pro to further reduce the cost of each individual sensor node.

To reduce the overhead of deployment, a software library was developed for the Jetson Nano to continuously receive thermal and RGB images from the FLIR One Pro, encrypt the images, and transmit the images securely to a cloud server for processing. Features of this library include parsing of raw data from the FLIR One Pro, encryption of the thermal and RGB images, and configuration files for quick setup for communicating with the server.

3.2 Environmental Sensing

From studies such as [6, 7], ambient temperature, and humidity are important features in estimating comfort. These features are also used in the ASHRAE 55 Standard for estimating thermal comfort. To incorporate these features in our system, we deploy environmental sensors consisting of a TMP35 sensor and a Huzzah Feather board to measure ambient temperature and humidity time series data, which is transmitted to the server every 15 seconds. Thermal measurements are associated with ambient temperature and humidity measurements based on the environmental sensors nearest to the comfort sensor node.

3.3 Thermal Comfort Feedback

Obtaining thermal comfort feedback from occupants is important, as it acts as a target for training comfort models, and provides insight into personal preferences to aid in personalizing comfort estimates. From the ASHRAE 55 Standard, occupants provide their perceived thermal comfort feedback on a 7-point thermal sensation scale. As shown in Figure 2, we provide a web interface where occupants can easily submit thermal comfort feedback. Thermal comfort feedback is submitted with an ID, which can be associated with identified thermal measurement data in Section 3.5.4.

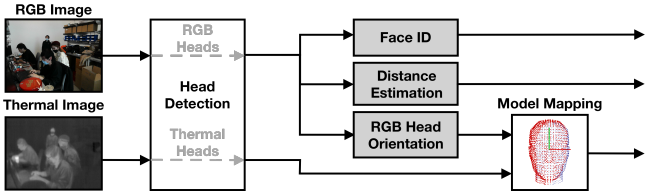

Figure 3: The feature extraction pipeline outputs participant ID, distance estimate, and facial thermal features for each detected occupant in the RGB and thermal images.

To increase the amount of training data that can be collected with less ground truth data (which requires human interaction), each comfort feedback response is used to label data in a small time window around the feedback response. We make the assumption that perceived thermal comfort does not change drastically for small time windows. Not only does this increase the amount of training data, but also provides multiple measurements to reduce effects of measurement error in single measurements.

3.4 Comfort Server

The comfort server is the central component of the system. The server is responsible for receiving RGB and thermal image streams, extracting thermal features and face identification information, and incorporating thermal comfort feedback and environmental sensing to learn comfort models. To receive data from the thermal comfort sensor nodes, Redis is used in a publish and subscribe pattern. RGB and thermal images can be continuously sent to independent streams set up for each sensor node. The server polls each stream via round robin scheduling, using the most recently published images from each stream in feature extraction. This design allows the flexibility to easily add additional sensor nodes.

RGB and thermal images from each stream are processed through a thermal feature extraction pipeline, described in Section 3.5. The pipeline outputs thermal comfort features, face identification and a distance estimate for each occupant detected in the images. This information is associated with relevant thermal comfort feedback and environmental sensing, and provided to the comfort models described in Section 3.6.

3.5 Thermal Feature Extraction

One of the most important features for personal comfort estimation is skin temperature measurements, which is useful for indicating the perceived comfort level of individual occupants. To extract these thermal features, the cloud server processes the RGB and thermal images through an image processing pipeline to extract the facial skin temperature features, distance, and the identity of each occupant in view of the sensor nodes.

As shown in Figure 3, there are five components in the thermal feature extraction pipeline: head detection, using a retrained YOLO v3 model; head orientation estimation, using a retrained state-of-the-art model; distance estimation; facial identification, using a custom convolutional neural network; and model mapping, using point clouds to produce a 3D thermal map for each occupant.

3.5.1 Head Detection. Studies such as [17] have utilized measurements from the hands and neck as features for predicting thermal comfort; however, these measurements are not always in view and can be difficult to detect. Detection of the head, on the other hand, provides more consistent measurement. Although many pre-trained object detection models are readily available, these models are not immediately applicable in our system for a variety of reasons. Firstly, models are often trained to detect faces, rather than heads; this is an important distinction, as faces that are not directly facing the camera result in lower detection accuracy. Secondly, models are not trained on datasets including people wearing face masks due to the onset of SARS-CoV-2. Finally, our thermal feature extraction pipeline requires head detection in thermal images as well, which is a domain not well explored in pre-trained models.

To overcome these limitations, we procured a custom training dataset consisting of two separate datasets. The first is a pre-existing head dataset from the South China University of Technology (SCUT) [16], which consists of over 100,000 labeled bounding boxes of heads from different perspectives and distances. The second is a hand labeled dataset from our own deployment in a commercial building, which provides data samples which are similar to our anticipated conditions, such as image quality and facial coverings. For this dataset, we hand labeled heads in 1000 RGB and thermal images each. After trying different state-of-the-art models, we chose to use YOLO v3 [18], as it meets the accuracy and latency requirements for our system. We retrained the network using the SCUT and hand labeled dataset, which outputs head bounding boxes in the RGB and thermal images, which are further processed for head orientation and facial identification.

3.5.2 Head Orientation Estimation. In [13], temperature features are extracted from the entire face, such as average or maximum temperature. However, temperature measurements vary for different parts of the face, which can result in major variations in temperature measurements depending on which parts of the face are visible to the thermal camera. Furthermore, face masks and other face obscurities can also change the measurements. One method to combat this source of error is to map temperature measurements to specific locations on the face, so that features can be extracted from more uniform surfaces (such as from the forehead region).

To determine the mapping from thermal image measurements to facial regions, we only require an estimate of the head orientation. With head orientation information, we can correctly orient a 3D head model, apply the measurements in the thermal image, and finally extract thermal features. To estimate head orientation, we retrain a state-of-the-art head orientation model, FSA-Net [27]. As is the case in head detection, a custom dataset is required to adapt the model to face masks and extreme angle perspectives.

We collected and hand labeled a dataset consisting of over 1000 head images with pitch, roll, and yaw angles. As discussed in Section 4.2, we found significant improvement in error after training with the hand labeled dataset.

3.5.3 Distance Estimation. As noted in [12, 13] the distance from the thermal camera affects surface temperature measurement. Rather than calibrate the temperature measurements based on distance, we include distance as an additional feature to comfort estimation. Our method to estimate the distance from the FLIR camera is based on the fact that the further the distance of an object from the camera, the smaller the object area in the image and vice versa. If we make

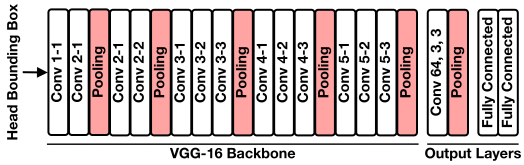

Figure 4: Face recognition convolutional neural network architecture, with VGG-16 backbone.

the assumption that a bounding box of a head projected in reality does not change significantly for perspectives or different people, a relationship between the area of the bounding box in the image (A), and the distance to the view plane containing the projected bounding box (D), can be derived as: $AD^2 = \text{constant}$.

We fit a numerical constant based on multiple ground truth measurements. From distance, the height and width of the view plane can be derived using horizontal and vertical field of view (HFOV and VFOV). Subsequently, the horizontal and vertical displacement (X_b and Y_b) in the view plane can be derived from the height and width of the view plane. Finally, the distance from the camera (R) can be estimated as $R^2 = X_b^2 + Y_b^2 + D^2$.

3.5.4 Face Identification. To associate comfort estimates to specific occupants, some method of identification is necessary. As our system already extracts bounding boxes of occupant heads, we can utilize a convolutional neural network to differentiate occupants and associate comfort estimates. Face recognition is a well studied field which has produced highly accurate models such as FaceNet and OpenFace [2]. As our deployment only requires a low latency and lightweight model for a select number of occupants, we developed a smaller VGG-16 based CNN. However, a state-of-the-art model can be substituted for larger numbers of occupants.

As shown in Figure 4, our network uses a VGG-16 backbone which feeds into a convolutional layer and dense layers for classifying occupants. To train the model to differentiate between occupants, head bounding boxes for each occupant are hand labeled using over 100 examples for each occupant. As described in Section 4.2, even with a low number of training samples, our system is able to differentiate well between occupants due to the low number of occupants and fixed positions of the cameras and workspaces. With larger numbers of occupants, larger datasets may be required to train a facial recognition classifier.

We note that privacy is a major concern, especially for camera based systems. There exist other methods for associating comfort estimates that are more privacy preserving, such as mobile device localization through WiFi or Bluetooth fingerprinting. One advantage of the sensor node design is the potential for moving privacy-sensitive operations, such as head detection, head orientation estimation, and face identification, onto the edge device. This is an important future work that can both preserve privacy as well as provide accurate comfort estimate association.

3.5.5 Model Mapping. The final step in feature extraction is the mapping of thermal measurements in the thermal image to a head model. At shown in Figure 5, this involves selecting the points that are visible to the camera based on the orientation estimate, flattening the points onto a plane, mapping the 2D thermal measurements to the flattened points, and reprojecting the flattened

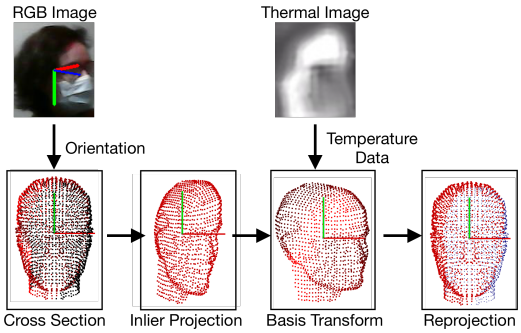

Figure 5: Mapping of thermal image data to a 3D point cloud head model.

points to recover the thermal measurement mapping. To manage the point cloud and project the thermal measurements, we utilize the C++ Point Cloud Library.

First, points that are visible to the camera are identified for mapping. To increase speed, this is approximated by bisecting the point cloud with a plane oriented by the orientation estimate. From the orientation estimation described in Section 3.5.2, we estimate a pitch, roll, and yaw for each detected head, which can be converted to a normal vector in Euclidean space $\vec{n} = \langle n_1, n_2, n_3 \rangle$. The equation of the plane that separates visible and invisible points to the camera is described as:

$$n_1(x - x_0) + n_2(y - y_0) + n_3(z - z_0) + D = 0$$

where $D = 0$ assuming the point cloud is centered at the origin. Note that a more precise selection of points can be obtained with methods such as raytracing; however, this increases computation time significantly. Once the points have been selected, the mapping between the 2D thermal image and the 3D points can be done in two steps. We first project the 3D points onto the prior defined 2D plane, and then change the basis of the plane into the cartesian coordinate space with a transformation matrix M. Let the point we want to project be defined as: $p = (x, y, z)$. then the projected point and matrix M can be computed as follows:

$$p_{proj} = p - \vec{n} * \vec{n} \cdot \vec{p}, \quad M = \begin{bmatrix} n_x^1 & n_x^2 & n_x^3 \\ n_y^1 & n_y^2 & n_y^3 \\ n_z^1 & n_z^2 & n_z^3 \end{bmatrix}$$

After the change of basis, the points will be aligned along one of the cartesian axes (x); thus, the thermal image can be fit onto the projected points by first finding the minimum and maximum of the other two axes, $y_{min}, z_{min}, y_{max}, z_{max}$. Each point corresponds to a pixel in the thermal image (p_x, p_y) as in the equation:

$$p_y = \frac{(y - min_y) * width}{max_y - min_y}, \quad p_z = \frac{(z - min_z) * height}{max_z - min_z}$$

Lastly, the thermal measurements can be associated to each point's original locations in 3D space to complete the model mapping. With a 3D representation of the thermal measurements, features can be extracted by computing statistics of groups of points based on their location on the 3D head model. For this system, we extract median, maximum, and average temperature features from six locations: forehead, left and right temple, left and right cheek, and nose areas. These features are calculated by selecting the points

on the head model that correspond to each region, and computing the statistics over the mapped temperature values of those points.

Note that until this point, the the system output is the 3D head model. The purpose of the pipeline is to provide an accurate temperature model to recover more precise facial temperature statistics.

3.6 Comfort Estimation

The goals of comfort estimation are to both provide comfort estimates for all occupants (a general comfort model), and to utilize human feedback to provide more accurate personalized comfort estimates (personal comfort models). There are two modes for comfort estimation: comfort training and comfort estimation. Initially, the system undergoes a training period involving collection of features and ground truth data (perceived comfort feedback).

One challenge that is present in our system is missing data. As described in Section 3.5.5, we extract features from six locations on the face, and some locations are hidden from view of the camera resulting in missing data in our dataset. To enable the use of this data, we utilize multiple imputation to fill in the missing values. This method enables us to use a significantly larger dataset for training, which aids in model accuracy.

Once data has been collected, a general comfort model is trained to provide comfort estimates even for occupants who have not given any comfort feedback, using all of the labeled data from all occupants. For more personalized comfort estimates, personal comfort models are trained for each occupant who has submitted comfort feedback. As thermal comfort preferences may vary between people, each model should be tuned to the specific preferences of the occupant. For these models, if a threshold value of comfort feedback has been received, only the labeled data for the occupant is used to train the personal comfort model.

4 EVALUATION

4.1 Experimental Setup

To evaluate our system, we deployed eight sensor nodes in eight distinct locations throughout two floors in the Northwest Corner Building at Columbia University to maximize spatial coverage. We obtained approval from the Columbia University Internal Review Board (IRB) to conduct this human research study. We collected data for ten occupants over the course of two weeks, with the first week as a training period and the second week as an evaluation period. For ground truth, we provided occupants with a web interface to input their perceived thermal comfort on the ASHRAE thermal sensation scale, as described in Section 3.3. In total, we collected over 2000 thermal measurements over two weeks.

4.2 Microbenchmarks

We first evaluate individual components of the feature extraction pipeline. We evaluated head detection by hand labeling bounding boxes of over 1000 heads, and compared with the output of the head detection network, as shown in Table 1. We achieved high precision, although the recall suggests that heads are not detected in all frames. This may be acceptable for our study, as multiple thermal measurements can be captured from each occupant. However, additional training data can improve the recall to provide even more robust comfort estimates.

Metric	Precision	Recall	RGB IOU	Thermal IOU
YOLO-v3	98.0%	91.0%	77.6%	80.3%

Table 1: Head detection precision, recall, and intersection over union on RGB and thermal images.

Metric	Pitch MAE	Yaw MAE	Roll MAE
FSA-Net (Retrained)	10.4°	6.4°	1.6°
Baseline FSA-Net	23.7°	12.5°	7.0°

Table 2: Head orientation estimation comparison between baseline and retrained FSA-Net models.

We evaluated head orientation estimation by hand labeling pitch, roll, and yaw of over 600 heads, and computing the mean average error (MAE) and root mean squared error (RMSE) of our retrained FSA-Net model and the baseline FSA-Net model as shown in Table 2. For all three angles, retraining reduced MAE, suggesting that retraining can help mitigate factors such as face coverings.

Metric	MAE	RMSE
Distance Estimation	11.2 cm	15.5 cm

Table 3: Distance estimation error.

For distance estimation, we computed MAE and RMSE for 500 detected heads. Table 3 suggests that while our method for estimating distance produces reasonable error, additional hardware (such as a depth camera) can be added to further reduce distance error. Finally, we found that our CNN with VGG-16 backbone was able to achieve a face recognition accuracy of 96.8%. This network works well in our system with a low number of occupants, but larger deployments may require more complex models.

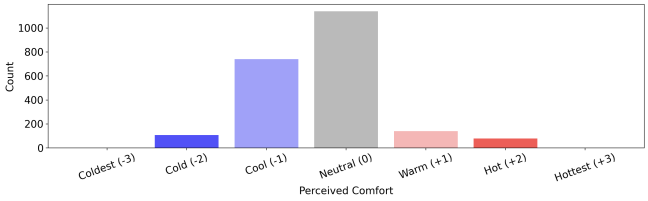

Figure 6: Histogram of thermal comfort measurements labeled with perceived thermal comfort feedback.

4.3 Comfort Estimation

To evaluate comfort estimation, we constructed training and evaluation datasets during the first and second weeks of the study by associating thermal comfort feedback with the collected thermal features. Models were trained on the training dataset and evaluated on the evaluation dataset. The ground truth data collected from ten occupants ranges from −2 to +2 on the 7-point thermal sensation scale and are shown as the histogram in Figure 6.

To evaluate comfort estimation of our system, we trained two different general comfort models: a random forest regressor with depth 4 (RF), and a linear regressor (LR). For the personal comfort models, we separated the dataset by occupant ID and trained a separate random forest regressor with depth 4 for each occupant. We also compare our system with two baseline models: the ASHRAE 55 Standard, and [13]. We calculate mean comfort votes to compare with the PMV model from ASHRAE 55. To replicate [13], we replace

the thermal features in the dataset with the maximum and average statistics, and train a linear regression model. The evaluation error on the second week data is shown in Figure 7.

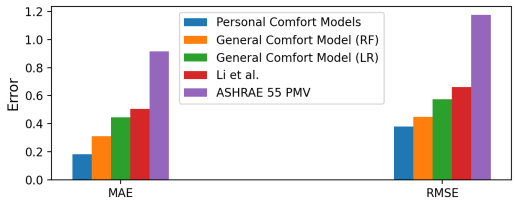

Figure 7: Comfort estimation error (PMV) for the general and personal comfort models, Li et al., and ASHRAE 55 PMV.

There are a few key takeaways from this study. Our system, as well as [13] achieves much lower MAE and RMSE than the ASHRAE 55 Standard, which suggests that directly measuring thermal features from occupants reduces thermal comfort estimation error significantly. Second, our system achieves lower MAE and RMSE than [13], suggesting that aggregating thermal features over the entire face may result in higher feature error. Finally, the personalized comfort models further reduce MAE and RMSE over the general comfort models. We observed up to a 2° F difference in what is considered cool thermal comfort between occupants, in similar environmental conditions. This suggests that individual comfort models can better capture variations in occupant thermal preferences. Note that we only measure temperature and humidity for the ASHRAE 55 baseline; future work can better capture improvements with additional measurements.

5 CONCLUSION

In this work, we describe the architecture, design, and implementation of a personalized comfort estimation system. The system extracts facial temperature features and is robust to changes in occupant facial perspectives to the camera. The system is scalable, allowing additional sensing nodes to be quickly deployed and integrated. We enable personalized comfort estimation by training individual comfort models, achieving lower comfort estimation error than general comfort models trained with data from all occupants. Over a two week study with ten occupants, our system reduced comfort estimation error by 64% over baseline methods.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation under Grant Numbers CNS-1704899, CNS-1815274, CNS-11943396, and CNS-1837022. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

REFERENCES

[1] Joseph G Allen, Piers MacNaughton, Jose Guillermo Cedeno Laurent, Skye S Flanigan, Erika Sita Eitland, and John D Spengler. 2015. Green buildings and health. *Current Environmental Health Reports* 2, 3 (2015), 250–258.

[2] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. 2016. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science* 6, 2 (2016).

[3] Ashrant Aryal and Burcin Becerik-Gerber. 2019. Skin temperature extraction using facial landmark detection and thermal imaging for comfort assessment. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 71–80.

[4] Yun Cheng, Kaifei Chen, Ben Zhang, Chieh-Jan Mike Liang, Xiaofan Jiang, and Feng Zhao. 2012. Accurate real-time occupant energy-footprinting in commercial buildings. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. 115–122.

[5] Joon-Ho Choi and Dongwoo Yeom. 2019. Development of the data-driven thermal satisfaction prediction model as a function of human physiological responses in a built environment. *Building and environment* 150 (2019), 206–218.

[6] Richard De Dear and Gail Schiller Brager. 1998. Developing an adaptive model of thermal comfort and preference. (1998).

[7] Poul O Fanger et al. 1970. Thermal comfort. Analysis and applications in environmental engineering. *Thermal comfort. Analysis and applications in environmental engineering*. (1970).

[8] Ali Ghahramani, Guillermo Castro, Simin Ahmadi Karvigh, and Burcin Becerik-Gerber. 2018. Towards unsupervised learning of thermal comfort using infrared thermography. *Applied Energy* 211 (2018), 41–49.

[9] Ji Jia, Chengtian Xu, Shijia Pan, Stephen Xia, Peter Wei, Hae Young Noh, Pei Zhang, and Xiaofan Jiang. 2018. Conductive thread-based textile sensor for continuous perspiration level monitoring. *Sensors* 18, 11 (2018), 3775.

[10] Caroline Karmann, Stefano Schiavon, and Edward Arens. 2018. Percentage of commercial buildings showing at least 80% occupant satisfied with their thermal comfort. (2018).

[11] Vahid Kazemi and Josephine Sullivan. 2014. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1867–1874.

[12] 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* 176 (2018), 246–261.

[13] 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 (2019), 113336.

[14] J Fergus Nicol and Michael A Humphreys. 2002. Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and buildings* 34, 6 (2002), 563–572.

[15] Haoran Ning, Zhaojun Wang, and Yuchen Ji. 2016. Thermal history and adaptation: does a long-term indoor thermal exposure impact human thermal adaptability? *Applied Energy* 183 (2016), 22–30.

[16] Dezhi Peng, Zikai Sun, Zirong Chen, Zirui Cai, Lele Xie, and Lianwen Jin. 2018. Detecting heads using feature refine net and cascaded multi-scale architecture. In *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2528–2533.

[17] 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*. 1212–1222.

[18] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[19] Marcel Schweiker, Gesche M Huebner, Boris RM Kingma, Rick Kramer, and Hannah Pallubinsky. 2018. Drivers of diversity in human thermal perception–A review for holistic comfort models. *Temperature* 5, 4 (2018), 308–342.

[20] 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* 16, 4 (2016), 420.

[21] ASHRAE Standard. 2010. Standard 55-2010, Thermal environmental conditions for human occupancy. *American Society of Heating, Refrigerating and Air Conditioning Engineers* (2010).

[22] Kwok Wai Tham and Henry Cahyadi Willem. 2010. Room air temperature affects occupants' physiology, perceptions and mental alertness. *Building and Environment* 45, 1 (2010), 40–44.

[23] Ahmet Uğursal and Charles H Culp. 2013. The effect of temperature, metabolic rate and dynamic localized airflow on thermal comfort. *Applied energy* 111 (2013), 64–73.

[24] Zhe Wang, Richard de Dear, Maohui Luo, Borong Lin, Yingdong He, Ali Ghahramani, and Yingxin Zhu. 2018. Individual difference in thermal comfort: A literature review. *Building and Environment* 138 (2018), 181–193.

[25] Peter Wei, Stephen Xia, Runfeng Chen, Jingyi Qian, Chong Li, and Xiaofan Jiang. 2020. A Deep-Reinforcement-Learning-Based Recommender System for Occupant-Driven Energy Optimization in Commercial Buildings. *IEEE Internet of Things Journal* 7, 7 (2020), 6402–6413.

[26] Peter Wei, Stephen Xia, and Xiaofan Jiang. 2018. Energy saving recommendations and user location modeling in commercial buildings. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. 3–11.

[27] Tsun-Yi Yang, Yi-Ting Chen, Yen-Yu Lin, and Yung-Yu Chuang. 2019. Fsa-net: Learning fine-grained structure aggregation for head pose estimation from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1087–1096.