A Low-Cost In-situ System for Continuous Multi-Person Fever Screening

Kaiyuan Hou*, Yanchen Liu*, Peter Wei*, Chenye Yang*, Hengjiu Kang*, Stephen Xia*, Teresa Spada†, Andrew Rundle*, and Xiaofan Jiang*

Columbia University* Columbia University Medical Center†

{kh3119, yl4189, pw2428, cy2540, hk3120, sx2194, agr3}@columbia.edu

ts126@cumc.columbia.edu jiang@ee.columbia.edu

2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN) | 978-1-6654-9624-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/IPSN54338.2022.00009

ABSTRACT

With the recent societal impact of COVID-19, companies and government agencies alike have turned to thermal camera based skin temperature sensing technology to help screen for fever. However, the cost and deployment restrictions limit the wide use of these thermal sensing technologies. In this work, we present *SIFTER*, a low-cost system based on a RGB-thermal camera for continuous fever screening of multiple people. This system detects and tracks heads in the RGB and thermal domains and constructs thermal heat map models for each tracked person, and classifies people as having or not having fever. *SIFTER* can obtain key temperature features of heads in-situ at a distance and produce fever screening predictions in real-time, significantly improving screening throughput while minimizing disruption to normal activities. In our clinic deployment, *SIFTER* measurement error is within 0.4°F at 2 meters and around 0.6°F at 3.5 meters. In comparison, most infrared thermal scanners on the market costing several thousand dollars have around 1°F measurement error measured within 0.5 meters.*SIFTER* can achieve 100% true positive rate with 22.5% false positive rate without requiring any human interaction, greatly outperforming our baseline [1], which sees a false positive rate of 78.5%.

Figure 1: ColumbiaDoctors - Midtown deployment — Left: nurse measuring a patient's temperature using infrared thermometer (used as ground truth); Right: *SIFTER* system running in the reception entrance (next to the nurse).

1 INTRODUCTION

Thermal cameras have recently gained exposure as a novel sensor capable of sensing temperature in a variety of applications. Originally developed for military applications, thermal cameras have been commercialized to impact a wider field of applications. Typically, these applications for thermal cameras do not require extremely precise measurements, such as for detecting gas and thermal leaks in buildings [2]. However, with the recent disruption of society due to COVID-19 [3], companies and government agencies have tried to leverage thermal cameras to measure accurate skin temperature for fever screening. While fever is not always a symptom of COVID-19, the ability to screen a subset of carriers from a large population is still extremely valuable. It can help prevent the spread of diseases such as COVID-19, or even seasonal flu. It is a valuable tool to help prevent future pandemics.

In a pandemic, fever screening systems are needed in a wide variety of public applications. Some applications require fever screening systems to handle high throughput, often with multiple entrances, including hospitals, universities, transportation hubs, and commercial buildings. Some applications requires a low cost system. In all of these applications, cost, accurate and high throughput as well as convenience are important considerations. Current fever screening systems that are designed for high throughput are not low cost, while a low cost system are not capable of high throughput.

There are two major types of fever detection systems: non-contact infrared thermometry devices (NCIT), and infrared thermography (IRT) systems. Studies such as [4, 5] have shown the potential of NCIT systems for fever screening. However, NCIT devices have certain drawbacks, most notably that an employee is required to be in close proximity to the patient. This increases risk of exposure to the employee and can require significant human effort. On the other hand, current IRT systems require specific criteria to be met for accurate temperature measurements, such as distance to camera and acclimation to indoor temperature. Furthermore, employees with specialized training are needed to operate and/or interpret results of the temperature measurements. More recently, research studies seeking to automate the screening of febrile humans [6–8] have encountered a number of challenges that limit full automation. These systems require the face to be fully shown and close to the camera no farther away than 10 to 50 cm [9].

The first challenge to better enable automatic screening of febrile humans is that the measurements should be robust to conditions that are not in the system's control. First, skin temperature is largely dependent on the ambient environment. For people who have not yet been acclimated to the indoor environment, temperature measurements will not be consistent [10]. Second, distance from thermal

978-1-6654-9624-7/22/$31.00 ©2022 IEEE
DOI 10.1109/IPSN54338.2022.00009

camera can can cause differences in thermal measurements. Third, measurements of a single person may change in the presence of multiple people. Although these conditions can sometimes be controlled through protocols, improving measurement robustness to these conditions brings advantages to the screening efficiency and enforcement of protocol.

Another significant challenge preventing automatic measurements is the accurate selection of temperature features. There are variations in skin temperature for different regions of the face, meaning that aggregate statistics will change depending on which regions of the face are visible to the thermal camera. Many current IRT systems utilize mean or maximum facial temperature, which can vary even for the same person depending on facial perspective or facial coverings (such as face masks, glasses, or hats). Other systems which detect facial landmarks may also have difficulties when certain facial features are covered. Furthermore, certain regions of the face are more predictive of actual body temperature, and thus accurate temperature feature extraction is critical.

In this work, we present *SIFTER*, a low-cost system based on a RGB-thermal camera for autonomous continuous fever screening of multiple people. *SIFTER* handles the above challenges in its design, and achieves the following key contributions:

(1) **Robust and in-situ temperature measurements:** We perform studies to reduce the effects of distance on temperature measurements, thus drastically increasing the operating distance range of the fever screening system. In addition to improving measurement accuracy at different ranges, *SIFTER* increases measurement throughput by removing the requirements for the subject to stand directly in front of the in-situ thermal camera for a prolonged period. This also reduces the disruption to people's normal activities, and helps to bring their lives back to normal. Furthermore, we studied the effects of cold outdoor temperatures on temperature measurements, and offer a method to flag these people for additional screening after acclimating to the indoor environment. In addition to protocols for enforcing acclimation, this allows systems to detect people who have not yet acclimated to indoor temperatures. Finally, we investigate the effects of multiple measurements in a single frame on temperature estimation accuracy.
(2) **Accurate Thermal Feature Selection:** Rather than rely on aggregate statistics to estimate temperature, we model thermal features by mapping temperatures onto a 3D head model. This enables two important ideas: first, detected thermal features are robust to changes in facial perspective, and second, facial temperature features can be selected directly from the 3D head model, without needing to train a network to specifically target regions of the face. We demonstrate improved accuracy over pre-trained models and advantages in precision and recall over baseline model [1].
(3) **Design for Ease of Deployment:** We designed the system to be low cost, easily deployable, and usable in both low and high throughput settings. *SIFTER* uses the 399 USD FLIR One Pro with a Jetson Nano, to keep the hardware cost less than 500 USD per unit. We developed a software library to deploy on the Jetson Nano, which enables automatic interfacing with the FLIR One Pro, encryption and transmission of RGB and thermal images, and configuration for cloud services. The cloud server is designed to provide temperature estimation and fever screening as a service for multiple clients, thus reducing direct cost of the system to the client.
(4) **Fever Screening Optimization:** We present variant gradient boosting method using asymmetric loss functions to target high temperature measurements in this fever screening application. We compared these variations against a number of other models and baseline methods, and demonstrate improvements in temperature estimation and fever screening. On a six month dataset collected from a multi-speciality medical practice, we achieved errors within 0.4°F at a distance of 2 meters and 0.6°F at a distance of 3.5 meters.

2 RELATED WORKS

Non-contact skin temperature sensing methods with thermal cameras are becoming more common for detecting diseases such as COVID-19. Researchers [1, 11, 12] have presented pipelines for autonomously measuring thermal features using thermal cameras. Further, many commercial companies have also developed thermal camera based solutions for fever detection, which is traditionally accomplished by thermometer, clinical records or genetic methods. However, there are a number of challenges which inhibit adoption of these temperature sensing systems, including the overall cost of the system and deployment restrictions.

Deployment cost is critical in many situations. Existing systems typically rely on higher end thermal sensors, to ensure some level of measurement quality. For example, thermal camera companies such as FLIR and Thermoteknix have begun developing systems for detecting fever in populations to combat the spread of COVID-19. These systems have been deployed in public places by government agencies, such as airports, as well as in workplaces by big companies such as Amazon. Wuhan Guide Infrared Co. sold out thousands of skin temperature screening systems in two weeks for deployment in airports, railway stations, etc, with price ranges from $\$5,000$ to $\$30,000$ [1]. However, for many scenarios such as workspaces, classrooms, and shared residences, it may be infeasible to deploy such expensive systems. In these cases, it is imperative that low-cost versions be able to perform quality measurements.

Minimizing the number of sensors is one method for reducing overall cost, and can be achieved by deploying the thermal camera at a distance to increase field of view. With a larger field of view, a thermal sensor is able to capture measurements from more people [13], thus reducing the need for additional sensors. However, the downside is the decrease in resolution; thus, it is important to have a system which can recover quality measurements at longer distances. Prior works have additional challenges in deployment restrictions: spatial camera restrictions and restrictions of subject behavior. For spatial restrictions, prior works typically require the camera to be placed in a location where people can easily be detected, such as in front of a desk [1] or at standing eye level for fever

[1] https://www.forbes.com/sites/jeremybogaisky/2020/02/10/we-are-running-as-fast-as-we-can-coronavirus-sparks-surge-in-demand-for-infrared-fever-detection-equipment/

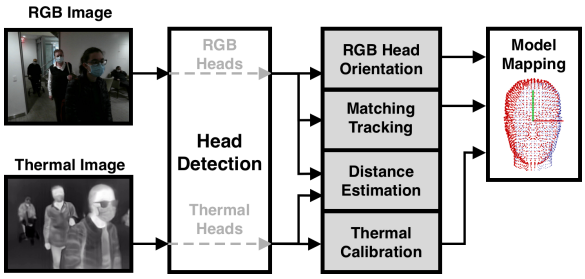


Figure 2: Block diagram of the fever screening pipeline.

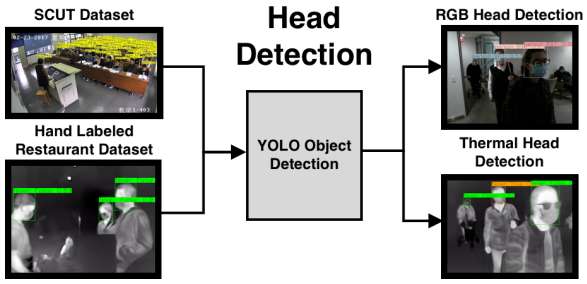


Figure 3: RGB and thermal datasets used for training the head detection network.

detection systems [2]. For many scenarios such as mass transit and classrooms, it may be infeasible to deploy such a setup; thus, a better system should have the flexibility of various spatial deployments.

As required by many prior systems, subject behavior is also controlled. In [1, 11], users are required to face forward to allow the thermal camera to easily detect facial features. In various skin temperature screening systems, users are required to stand in front of the thermal camera. Controlling subject behavior is unnatural, and will not only increase the time and labor costs but also limit the application of the system. Thus, a better system should be able to continuously screen multiple people in view and in-situ, without active participation from human subjects.

To address these issues, we developed a system which utilizes smaller numbers of low cost sensors, can run continuously and in different environmental conditions, and can screen multiple people simultaneously without requiring active participation. This system detects and tracks multi-person heads, estimates head orientation, reduces thermal measurement errors, and maps facial thermal features at-scale to 3D point-cloud head model to enable fever screening. Because *SIFTER* is low-cost, easily installed, non-intrusive, and robust in a variety of conditions, *SIFTER* is a critical step towards realizing systems and applications to create healthier city [14–27] and building environments [28–35].

3 FEVER SCREENING PIPELINE

The goal of this fever screening pipeline is to estimate skin temperature features for multiple occupants, given a stream of RGB and thermal images that are noisy, low-resolution, uncalibrated and without opportunistic observations of partial facial features, to classify febrile humans. *SIFTER*'s fever screening pipeline consists of six blocks, including: head detection, orientation regression, distance estimation, emissivity correction, matching and tracking, and model mapping. Initially, the RGB and thermal images are passed through an object detection pipeline trained to detect heads of all people in view. From the detected RGB and thermal heads, we estimate orientation, correct for emissivity, and estimate distance from the sensor. The thermal and RGB heads are then matched and tracked with previous frames. Orientation and thermal data is used to map skin temperature features onto a 3D point-cloud head model. An illustration of the different components of *SIFTER*'s fever screening pipeline is shown in Figure 2. The hardware used in this system to capture RGB and thermal images is a common low-cost RGB-thermal camera, FLIR ONE Pro.

3.1 Head Detection

To our knowledge, there are no open source networks trained specifically to detect faces or heads in thermal images. For *SIFTER*, one requirement is the detection of heads in both the thermal and RGB images. To verify that existing networks [36–38] are unable to detect faces or heads in RGB and thermal images, we performed an empirical study using YOLOv3 [39], a popular detection network. Our study showed that the pre-trained network does not perform well on RGB or thermal images. For RGB images, networks are trained on a large number of classes, to improve versatility. However, there is no class which includes the front and back of heads; thus, YOLO is not pre-trained to detect heads from the back. Further, standard datasets [40] do not include thermal images, which not only removes color information, but also removes landmarks or features important for identifying faces. Finally, due to the current pandemic, most people are wearing face masks, which is not included in popular image datasets, reducing the accuracy of pre-trained models even further.

To address this issue, we explored two options: pre-existing datasets, and a custom labeled dataset. For pre-existing datasets, we utilize the head dataset from the South China University of Technology (SCUT) [41], consisting of over 100,000 labeled bounding boxes of heads from different directions. This dataset was chosen over another head dataset [42] due to similarities in environmental conditions. The SCUT dataset is taken inside classrooms in a university, which is similar to the indoor settings in our application.

Secondly, we considered a custom labeled dataset, which carries two advantages. For RGB images, we can provide examples which are similar to our anticipated conditions, such as image quality and facial coverings. In addition to RGB images, we also chose to incorporate labeled thermal images, as this would help a network recognize heads in the thermal domain. We deployed *SIFTER* in a local restaurant to collect over 100 hours of images, and hand labeled 1000 images each of RGB and thermal images. More information about this deployment is described in Section 5.1.

As a baseline for head detection, we explored image processing methods for identifying heads, especially in the thermal domain. One possibility for identifying heads in the thermal domain is by identifying bodies with high temperatures compared to the background. However, there are possibilities for high temperature

[2]https://www.reuters.com/article/us-health-coronavirus-amazon-com-cameras/exclusive-amazon-deploys-thermal-cameras-at-warehouses-to-scan-for-fevers-faster-idUSKBN2200HT

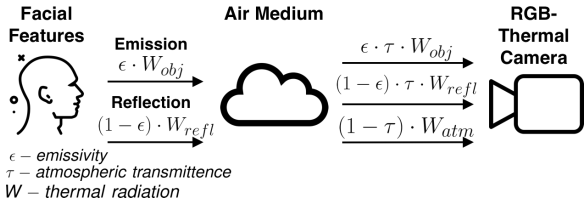


Figure 4: Model of thermal radiation captured by the sensor.

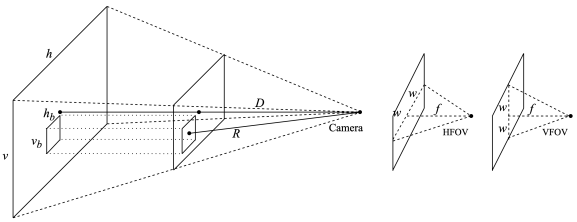


Figure 5: Distance estimation method using the size of the bounding box to estimate distance.

sources which are part of the environment, which may result in a high false positive rate.

From our own tests, we determined that YOLO meets the requirements for accuracy and has low latency. We trained YOLO using the SCUT head dataset as well as the custom RGB and thermal image dataset in different strategies, Figure 3. After training, the detection network outputs bounding boxes of heads in the RGB and thermal images, which are further processed for head orientation and thermal calibration.

3.2 True Temperature Estimation

The main sensor in common thermal cameras, as well as in the FLIR ONE Pro, is a microbolometer focal plane array, consisting of a number of pixels which sense radiation. Each pixel, similar to visible light camera pixels, senses radiation in a conical area visible through the lens of the camera. Typically, the sensor is calibrated after manufacturing to eliminate sources of error due to pixel variance. However, the sensors are calibrated against a blackbody with known radiance; most surfaces in the real world are not close to blackbody radiators. These sensors are prone to multiple sources of error due to the differences between real world environments and laboratory settings. For our system, there are two primary sources of error that we seek to minimize: surface emissivity, and spot-size effect. Reducing the effects of these phenomena is critical to robustly estimating the true temperature.

3.2.1 ***Emissivity.*** The radiation energy measured by the sensor is not equivalent to the radiation energy given off by the measured surface. The first reason is the *emissivity* of the surface; most surfaces do not emit 100% of the radiation energy, but instead also reflect energy from the surrounding environment. The second reason is due to transmittance; the medium which the radiation travels through can lead to attenuation, while introducing radiation energy from the atmosphere. Thus, it is critical to correct for these phenomena to recover a more accurate measurement.

Figure 4 illustrates the transmission of radiation energy from a surface to the sensor. Radiation energy emitted from the surface consists of a linear combination of the energy of the body W_{obj}, and the energy of the reflection W_{refl} (typically surrounding surfaces). The critical parameter is emissivity of the surface ϵ, which for human skin is between 0.95 and 0.98.

After the radiation energy leaves the surface, atmospheric energy W_{atm} is introduced as a linear combination with parameter τ. This parameter is the atmospheric transmittance, which is dependent on a number of factors including the distance between the surface and the sensor, and environmental conditions such as the humidity. The total radiation energy received by the sensor is thus:

$$W_{tot} = \epsilon\tau W_{obj} + (1-\epsilon)\tau W_{refl} + (1-\tau)W_{atm}$$

Temperature can be related to radiation energy by the Stefan-Boltzmann law: $W = \sigma T^4$, where σ is the Stefan-Boltzmann constant. Previous works [43] have shown that humidity and temperature change atmospheric transmittance heavily only when the distance between the body and sensor is on orders of hundreds of meters. However, our system is intended for distances of less than 5 meters, where distance has a much larger impact than humidity and temperature. After measuring the ambient temperature and estimating the atmospheric transmittance according to [43], we can recover the original temperature T_{obj} of the surface:

$$T_{obj} = \sqrt[4]{\frac{W_{tot} - (1-\epsilon)\tau\sigma T_{refl}^4 - (1-\tau)\sigma T_{atm}^4}{\epsilon\tau\sigma}}$$

3.2.2 ***Distance Estimation.*** A thermographic camera measures the temperature of objects by sensing infrared radiation. The spot-size effect influences the measured radiation; the measured temperature of human skin changes with distance from the thermographic camera. Because the FLIR One Pro camera doesn't measure depth, we require a method to estimate distance from captured RGB and thermal images.

Our method to estimate the distance from the FLIR camera is primarily based on perspective projection (the observation that the further away the object is from the camera, the smaller the object area is in the image). Figure 5 shows the area-based estimation method, assuming that the lens is a small hole. The definition of horizontal field of view (HFOV) and vertical field of view (VFOV) for an RGB lens is HFOV = $2\arctan(w/f)$ and VFOV = $2\arctan(w/f)$, where w refers to the dimension and f is the focal length. The two large rectangles are the view plane of the camera at different distances, and the small rectangle in the view plane is the bounding box of one detected head. According to the proportional relationship between the actual distance and the pixel distance, and the definition of HFOV and VFOV:

$$H/2 = w_H = D \times \tan(\text{HFOV}/2),\ H_b = 2w_H \times h_b/h$$
$$V/2 = w_V = D \times \tan(\text{VFOV}/2),\ V_b = 2w_V \times v_b/v$$

The actual area size of object is constant: $S_b = V_b H_b = \text{constant}$. Then we can derive the following formula, and roughly estimate D, the distance from camera to view plane, once the average constant is trained through multiple ground truth measurements:

$$v_b h_b D^2 = \text{constant}$$

The distance R, distance from camera to object, is calculated through Pythagorean Theorem, where x_b, y_b are the central coordinates of bounding box:

$$R^2 = X_b^2 + Y_b^2 + D^2 = (\frac{x_b}{h} \times H)^2 + (\frac{y_b}{v} \times V)^2 + D^2$$

To minimize bias in the RGB domain, we also compute the distance R from the thermal domain and apply stereo vision [44] to calculate the final distance.

3.3 Head Orientation

The RGB and thermal images capture only a certain perspective of each head, which can cause issues for landmark detection networks, and cascaded CNNs [45]. Since the ultimate goal is to use thermal values of specific areas of the face to perform fever detection, a correspondence between the pixel values and parts of the face is required. This correspondence can be determined by utilizing the head orientation. Specifically, the pitch, roll, and yaw of each head is sufficient to determine the correspondence. We start with a neural network implementation FSA-Net [46]. FSA-Net takes a single image as input, predicts head poses based on regression and feature aggregation, and output pitch, roll and yaw of each head in the image.

The main challenge with integrating this network into this system is three-fold. First, many of the detected heads are not directly facing the camera. The original FSA-Net trained on the 300W-LP dataset [47] only estimates orientations within 100 degrees (reference is straight forward), meaning that some images of the side head and most of the back head will not produce an estimate. However, side views of a face can still provide important thermal information. Second, due to current circumstances, many people are wearing masks, which are not included in the original training sets for FSA-Net. Lastly, there are differences in the image quality between the original training sets and the FLIR One camera, which can cause a loss in orientation accuracy.

To increase the orientation estimation capability of the FSA-Net, we hand labeled a custom dataset of head images at various pitch and yaw angles. Similar to the head detection network, we hand labeled 1000 images of heads with orientation taken from the restaurant deployment of *SIFTER*, which gives a variety of orientation angles, many examples of people wearing facemasks, and similar image quality. After retraining FSA-Net with this new dataset, we noted an increase in orientation accuracy especially for people wearing masks, and for angles outside the original 100 degree field. We also noted that accuracy suffers for angles where the majority of the face is covered by hair; this is potentially due to the fact that key identifying features for estimating orientation (such as the nose, eyes) are not present for angles towards the back of the head.

3.4 Matching and Tracking

Once head orientation has been extracted from RGB images, and thermal correction has been done on thermal images, it is important to *match* temperatures from the thermal images onto the corresponding location on the face in the RGB image. Furthermore, this information should be *tracked* across successive frames to construct time dependent information, which may be important for reducing error and potential future applications.

The difficulty in matching is that due to perspective changes between the thermal and RGB cameras, the same person will not appear in the same location in both images. Further, in the case of the FLIR ONE Pro camera, the field of view and image resolution are different between the two cameras. To successfully match heads in the RGB and thermal images, we first calibrate the images in both domains to correct the distortion. After calibration, we compare the number of detected heads in two images. If the numbers are equal, RGB heads and thermal heads are corresponded in order from left to right on the image. Because the lenses are facing the same direction and the relative positional relationships of people are same. However, if the numbers are not equal, there is a missed or wrongly detected head. We use Euclidean distance from the bounding box centroids to determine the closest match and remove the unmatched heads. Although there is significant potential to improve this matching algorithm, this is out of the scope of this work. In practice, this matching method achieves 94% accuracy rate.

The current framerate of our system is limited due to the transfer of images to the cloud for analysis. Due to this limitation, popular tracking algorithms that rely on low object displacement such as optical flow are not reliable. For our application, we implemented a Kalman filter based on the centroids of the bounding boxes in each frame. This method allows tracking of heads in the RGB domain across multiple frames. Although there are more complex methods for tracking that do not rely on high framerate and may be more reliable (such as particle filtering), in practice, the Kalman filter provides enough accuracy in tracking for our application.

Lastly, we combine matching and tracking. For each pair of RGB and thermal images, the bounding boxes of heads are matched. In consecutive frames, a Kalman filter is used to track the centroids of bounding boxes in the RGB domain only. The one set of matched and tracked bounding boxes corresponds to a single person in multiple frames, and is passed on to model mapping to produce a 3D thermal model.

3.5 Mapping Features to 3D Head Model

Before thermal features can be extracted for skin temperature estimation, we need to determine a mapping of the thermal image to a facial model. For instance, if we want to use the left cheek region in the fever classification model, we need to track the pixels that correspond to the left cheek in the thermal image.

We can utilize the thermal image with the orientation information to project values onto a head model. For the head model, we chose a standard 3D point cloud model, which we will project our thermal values onto. The computation for keeping track of the thermal values is done in the C++ Point Cloud Library.

To map the thermal values from the thermal image to the point cloud model, we first need to determine the points on the head model which are visible from the camera. The orientation predicted by the orientation network in Section 3.3 provides the angle difference between the orientation of the head with respect to the camera. This angle difference can be converted to a normal vector of the form $\vec{n} = \langle n_1, n_2, n_3 \rangle$.

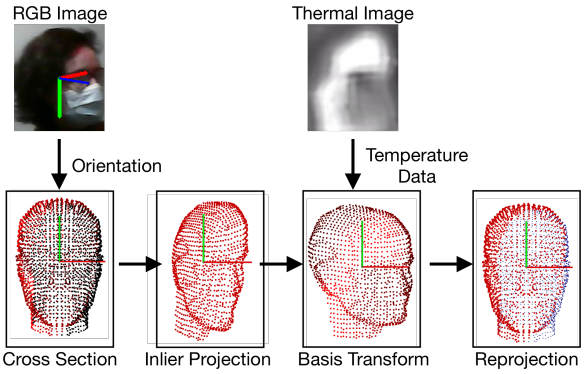


Figure 6: Mapping thermal features onto a 3D point cloud.

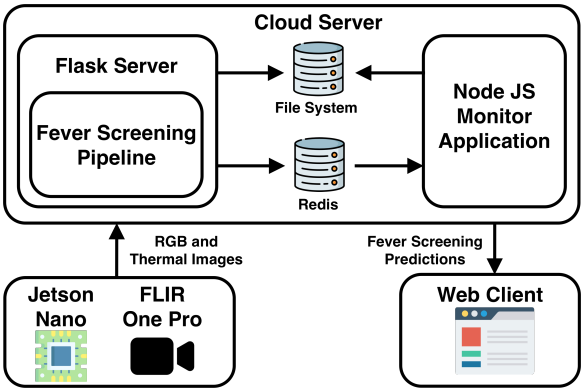


Figure 7: Block diagram of *SIFTER* architecture.

To increase speed, the point cloud is bisected along a plane intersecting the center of the point cloud. The equation of the plane can be defined 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. Only the points above the plane are considered to be visible by the camera. A more refined selection of points can be done by raytracing from the camera location; however, this increases the 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 ways. We first project the 3D points onto the prior defined 2D plane. Let the point we want to project be defined as: $p = (x, y, z)$, then the projected point can be computed as follows:

$$p_{proj} = p - \vec{n} * \vec{n} \cdot \vec{p}$$

A change of basis changes this plane into the Cartesian coordinate space by using the transformation matrix M:

$$M = \begin{bmatrix} n_x^1 & n_x^2 & n_x^3 \\ n_y^1 & n_y^2 & n_y^3 \\ n_z^1 & n_z^2 & n_z^3 \end{bmatrix}$$

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:

$$p_y = \frac{(y - min_y) * width}{max_y - min_y}$$

$$p_z = \frac{(z - min_z) * height}{max_z - min_z}$$

In addition, to help reduce noise in the pixel values, we combine observations with previous measurements. The temperature estimates for each point on the head model is saved, and new measurements are averaged to produce new temperature estimates. The overview of thermal feature mapping is shown in Figure 6.

3.6 Fever Detection

With constructed 3D head models, we can choose which regions of the face to use as features for estimating temperature and fever screening. The area around the temples and the forehead can be measured to estimate body temperature and detect fever [48, 49]. Furthermore, we utilize a temporal thermometer to collect ground truth which measures forehead and temple skin temperature. For these reasons, we extract temperature statistics from three regions: left temple, right temple, and forehead. We calculate average, median and maximum temperature values for these three regions as the thermal features for temperature estimation and fever screening.

4 SYSTEM ARCHITECTURE

We designed the *SIFTER* with three components: a sensor node, a cloud server, and a client, as shown in Figure 7.

4.1 Sensor Node

The sensor node is responsible for capturing RGB-thermal images, encrypting the images, and transmitting the images to the cloud server for processing. Once the system has been deployed, the sensor node is able to run continuously without human intervention to prevent any potential exchange of bacteria or viruses. There are two criteria which are emphasized to improve deployability in different environments: cost and configuration.

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.

4.2 Cloud Server

After image data is transmitted from the sensor node to the cloud server, the cloud server is responsible for decrypting and decoding the image, running the processing pipeline, saving results in the file system or database, and serving the web client. One major benefit of the cloud server is the ability to provide temperature estimation and fever screening from RGB and thermal images as a *service*, which can reduce the cost to the client significantly.

The cloud server consists of a Flask back-end server, a Redis in-memory database and a Node.js web server. Flask is a python library that can handle image data sent from sensor node. This flask back-end server handles the image data and calls the image processing pipeline. The Node.js web server serves the web client. Image data and detection results are shared between the flask back-end server and web client using the Redis in-memory database. Our cloud infrastructure can be deployed on a typical server or through common hosting services, like Amazon AWS. In the restaurant deployment, we deployed a cloud server in a desktop computer, with an Intel i5 CPU, 8G RAM and Nvidia TitanX GPU. In the medical practice deployment, a cloud server is deployed in a Google Cloud Platform (GCP) Compute Engine instance, with 4 virtual CPU cores, 32G RAM and an Nvidia P100 GPU.

4.3 Web Client

The web client is a website opened on another device that enables clients, such as hospital employees, to monitor images collected from the sensor node and the fever screening results from the system pipeline in real-time. The web client consists of a Node.js web server and a client side. The Node.js web server is part of the cloud server and is hosted in the GCP Compute Engine instance, in which the Node.js is a open source javascript run-time environment. On the client side a web page is served giving clients the capability to monitor real-time RGB-thermal images, and historical images. Hospital employees can also delete both raw images and detection results through the interface if requested by a patient, in accordance to the terms of the IRB for this project.

5 MICRO-BENCHMARKS

We first evaluate the individual components of the fever screening pipeline to demonstrate the system's ability to accurately select and extract thermal features. An early prototype of *SIFTER* was deployed in a local restaurant to gather images for training the various models in the fever screening pipeline.

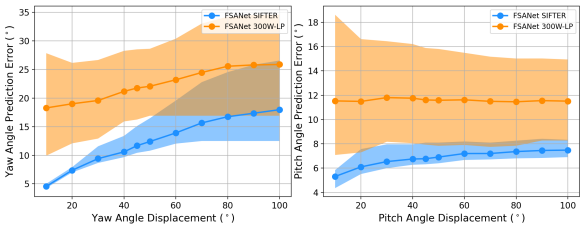
5.1 Local Restaurant

To gather preliminary images for training, an early prototype was deployed in a local restaurant. The system ran continuously for 3 weeks, more than 30 000 images have been gathered. A portion of these images were labeled to train and evaluate early pipeline models in *SIFTER*, including the YOLO head detection model and FSA-Net head orientation detection model.

5.2 Head Detection

As described in Section 3.1, we primarily focus on detecting heads by training YOLO [39] on different datasets, namely the SCUT head dataset [41] and a custom labeled RGB and thermal dataset taken from images at the local restaurant deployment. By incorporating both RGB and thermal images in the training data, we allow YOLO to learn head features from both image domains. Datasets were balanced to not provide bias towards either domain.

We compared YOLO trained with different datasets. The network was first trained with the SCUT head dataset only, to determine precision, recall, and intersection over union (IOU) for RGB images. Next, we trained YOLO with interleaved RGB and thermal images to improve performance on detecting heads in the thermal domain. When trained with both types of images, YOLO was found to perform better on both RGB and thermal images in precision, recall, and IOU, as shown in Table 1. Note that all models trained and tested in Table 1 include images from the SCUT dataset and the restaurant dataset. We see that there is relatively little change in performance in the RGB domain. However, there is a significant improvement in recall in the thermal domain, due to the addition of thermal images in the training set.



(a) Yaw angle. **(b) Pitch angle.**

Figure 8: Average yaw and pitch angle prediction error at different yaw and pitch angle displacements.

5.3 Head Orientation

As described in Section 3.3, the pre-trained FSA-Net implementation did not meet accuracy requirements. One major change is the inclusion of face masks, which greatly affects features that are used by FSA-Net to estimate orientation. We hand labeled orientation for over 500 head bounding boxes from the restaurant dataset, and retrained and evaluated FSA-Net. As shown in Figure 8, the retrained FSA-Net significantly reduces error in yaw and pitch angles. Note that the last axis, roll, is not shown as this angle varies the least, and the pre-trained version already achieves a low prediction error.

5.4 Distance Estimation

As described in Section 3.2.2, we estimate distance from the area of the bounding boxes in the RGB and thermal images. For training

Test on RGB Set	Precision	Recall	IOU
YOLO with RGB	97.0%	92.1%	72.8%
YOLO with RGB-Thermal	99.0%	98.0%	80.3%
Test on Thermal Set	**Precision**	**Recall**	**IOU**
YOLO with RGB	99.8%	8.7%	18.8%
YOLO with RGB-Thermal	98.0%	99.0%	83.9%

Table 1: Precision, recall, and IOU on RGB and thermal test sets when training on different training sets.

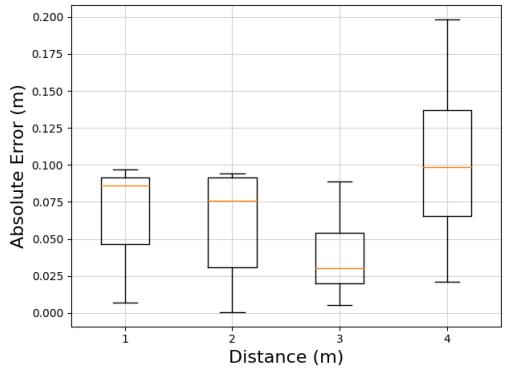


Figure 9: Absolute distance estimation error of the described area method on the dataset at different distances.

data, we labeled over 500 heads with ground truth distances. After training, we evaluated the method on 25% of the dataset, as shown in Figure 9. With bounding boxes up to 3.5 meters away, our method achieves an average accuracy of 0.064 meters. Intuitively, the error and variance increases at larger distances. More advanced methods can be utilized in future works to further reduce distance error.

5.5 Matching and Tracking

To evaluate matching and tracking accuracy, we selected 500 frames from the restaurant dataset and labeled bounding boxes with a unique ID, which associates bounding boxes across successive RGB and thermal frames. This simple methods can achieve a 94.4% accuracy in mathcing and 91.3% accuracy in tracking, respectively. Although these methods are acceptable for the restaurant and medical practice deployments (with few people in each frame), we recognize that more complex environments (especially greater numbers of people in frame) will require more complex methods to improve both matching and tracking accuracy.

6 REAL-WORLD EVALUATION

To evaluate the temperature estimation and fever screening capabilities of *SIFTER*, we conducted a real world user study to compare *SIFTER* with the clinically validated Withings Thermo NCIT[3]. We received IRB approval to deploy *SIFTER* on the ground floor at ColumbiaDoctors - Midtown. Over the course of six months, we collected a dataset consisting of measurements from over 4000 people with *SIFTER* and the Withings Thermo NCIT.

6.1 Medical Practice Deployment Dataset

In the medical practice deployment at ColumbiaDoctors - Midtown, patients enter through a single entrance either from the outside, or from a different floor in the medical practice. The patients approach the reception desk, which enables *SIFTER* to record multiple measurements at different distances, as shown in Figure 10. An onsite nurse records the ground truth temperature with the Withings Thermo NCIT shown in Figure 1, which is not in view of *SIFTER*. Our dataset consists of the following features:

[3]Withings Thermo: https://www.withings.com/us/en/thermo

(1) ***SIFTER* Temperature Features.** Average, median and maximum temperatures are recorded for the forehead, left temple, and right temple regions for each detected person. These features are extracted from the 3D head model described in Section 3.5.
(2) **Baseline Temperature Features.** For comparison, we also collected baseline temperature features commonly used by other fever screening systems, including maximum and average temperature of the entire face.
(3) **Distance.** As described in Section 3.2.2, distance for each detected person is estimated and stored as a feature.
(4) **Acclimation.** Whether a person is acclimated to the indoor environment also affects skin temperature. In this deployment, a person can enter the ground floor practice from either the outside or from a different floor. Both of these routes are visible to *SIFTER*, so an acclimation feature can be assigned to each person.
(5) **Delay.** For people who are not yet acclimated to the indoor environment, the skin temperature may change between successive measurements. To adjust for this temporal change, we add a delay feature, which is the time between the *SIFTER* measurements and the ground truth measurement.
(6) **Outside Temperature.** The effects of temperature acclimation on skin temperature depend on both the outside and inside environments. We also correspond each temperature measurement with the current outside temperature to account for this effect.
(7) **Ground Truth.** An onsite nurse measures ground truth temperature using the Withings Thermo NCIT. This measurement is saved along with a timestamp, which is associated with the *SIFTER* measurements. This measurement acts as the label for machine learning regression.

6.2 Fever Screening Study

To evaluate the potential screening capabilities of *SIFTER*, we implemented and trained a number of different models to measure fever screening performance. We explored two categories of models:

(1) ***SIFTER* Regression Models:** We implement and train linear regression, random forest and gradient boosting models using *SIFTER* temperature features to predict ground truth temperature.
(2) **Baseline Models:** We also implement baseline models, average, maximum, and Li et al. [1] to compare against the *SIFTER* models. These models use the baseline features, such as average and maximum facial temperature, rather than the features extracted from the 3D head models. For Li et al., we train a linear regression model on the average and maximum facial temperatures. Li et al. is an important baseline as it represents prior works that utilize statistics on the entire facial region, rather than on specific facial features.

In addition to the standard regression models, we also explored the performance of a few custom gradient boosting models. While training regression models such as random forest with depth 10, we noticed through inspection a general underestimation of predicted temperature for higher ground truth temperature values. While this is an undesirable characteristic, it also provides an opportunity to

Figure 10: Real-time images available to hospital staff displaying detected people and fever screening estimates.

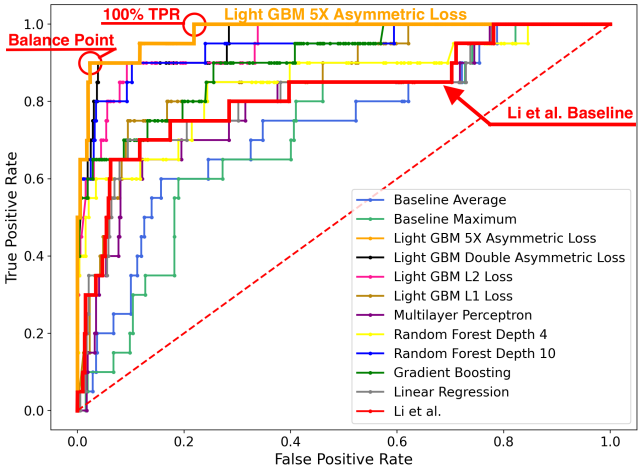
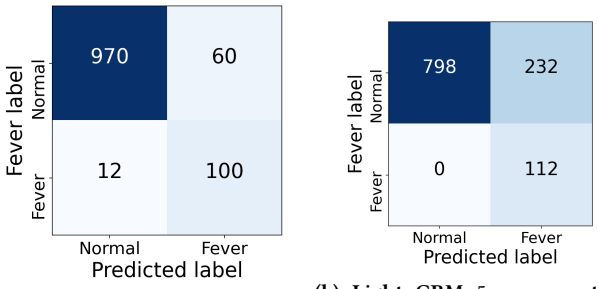


Figure 11: ROC curves for different regression models.



(a) Light GBM 5x asymmetric loss with balance point threshold.

(b) Light GBM 5x asymmetric loss with 100% true positive rate (recall) threshold.

Figure 12: Fever prediction confusion matrices.

potentially improve febrile human detection. The intuition is that measurements with higher ground truth temperatures are closer to the decision boundary, and thus having more accurate estimations for these temperatures will improve the decision boundary.

To improve estimation of the high ground truth temperature measurements, we designed asymmetric loss functions to give a greater penalty for underestimation of predicted temperature. We found empirically that asymmetric L2 loss improved ability to discriminate febrile humans. To implement asymmetric loss, we utilize LightGBM, a gradient boosting library which allows definition of loss. We define the asymmetric L2 loss, where $f(x)$ is the ground truth and $\hat{f}(x)$ is the prediction as follows:

$$L_i = \begin{cases} \alpha(\hat{f}(x) - f(x))^2 & f(x) - \hat{f}(x) \leq 0 \\ (\hat{f}(x) - f(x))^2 & f(x) - \hat{f}(x) > 0 \end{cases}$$

α is a tunable parameter that defines what factor we asymmetrically penalize temperature underestimation. We found that the mean temperature estimates are closer to ground truth for high ground truth temperatures for Light GBM with $2x$ and $5x$ penalty than with Random Forest with Depth 10.

To evaluate fever screening, we modified the label of the medical practice deployment dataset. The Center for Disease Control and Prevention (CDC) threshold for fever is 100.4°F. However, we choose a more conservative threshold: 98.6°F to minimize the false negative rate to ensure that close to all people who actually have fevers are detected. We labelled the ground truth measurements above 98.6°F with a 1 to indicate fever, and measurements at or below 98.6°F with a 0 (indicating normal temperature). For overall evaluation, we remove occupants who are not acclimated to the indoor environment, as this has a substantial impact on the accuracy of the fever screening system. We present and discuss possibilities for detecting non-acclimated occupants in Section 6.2.3.

For this evaluation, each model was trained and evaluated on a 75% : 25% train and test split on the medical practice deployment dataset. We compared all of the described models on a receiver operating characteristic curve (ROC curve) by sweeping the threshold temperature to understand the screening capabilities of each model. As shown in Figure 11, the best performing models are the Light GBM models with asymmetric loss functions, followed by Light GBM with symmetric L2 loss and Random Forest with depth 10.

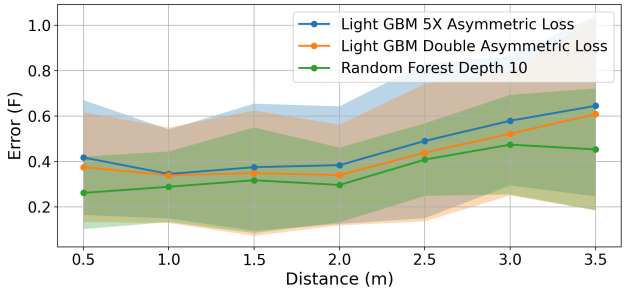


Figure 13: Temperature estimation error comparison between Random Forest (Depth 10), Light GBM with double and 5x asymmetric loss for different measurement distances.

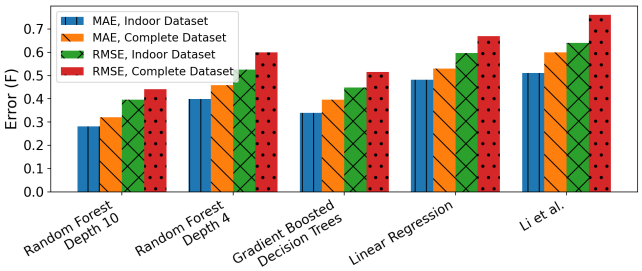


Figure 14: Mean Absolute Error of regression and baseline model temperature predictions on datasets including and excluding people not acclimated to the indoor environment.

This suggests that by focusing on improving temperature estimation accuracy for high ground truth temperature data, we are able to improve the ability of these models to screen for febrile humans.

In Figures 12a and 12b, we show the confusion matrix for fever screening using Light GBM with 5x asymmetric penalty and thresholds at the balance point and at 100% recall. If we select the balance point, or *max TPR − FPR* where TPR and FPR are the true positive rate and false positive rate respectively, we are able to achieve an 89.3% recall with 62.5% precision for 5x asymmetric penalty. To reduce the spread of virus, we aim to detect all febrile humans, even if we need to sacrifice the false positive rate. We can operate *SIFTER* at different operating points, and can optimize for 100% TPR. At 100% TPR, meaning we correctly classify all potentially febrile people, our system can achieve 22.5% FPR using Light GBM with 5x asymmetric penalty. In comparison, the baseline methods can only achieve 100% TPR by greatly sacrificing FPR. The baseline average, maximum and Li et al. achieve FPR of 79%, 81%, and 78.5%. Although a 22.5% FPR may seem high, we note that people who are detected to have a fever, will likely undergo further screening by health professionals or more accurate systems to confirm our measurement. Although this requires another stage of screening, our system still has the potential to significantly reduce human intervention for the majority of people without symptoms.

6.2.1 Distance Performance. In addition to the overall evaluation, we also studied characteristics of the medical practice deployment dataset. For this purpose, we created five different datasets each consisting of a subset of the main dataset: *Whole Dataset*, consisting of the entire dataset; *Indoor Dataset*, consisting of only people acclimated to the indoor environment; *Outdoor Dataset*, consisting of only people not acclimated to the indoor environment; *Single Dataset*, consisting of people who are measured independently; *Multiple Dataset*, consisting of people who are measured with other people in view of the sensor. We also divided the dataset by distance ranges to study the effects of distance on measurement accuracy.

Since the distance from the RGB-thermal camera has a significant effect on the measured temperatures, it is critical to evaluate the effect of distance on the predicted temperature accuracy. We tested three of the best performing regression models on varying distances, Random Forest with depth 10, Light GBM with double asymmetric loss and with 5x asymmetric loss. As shown in Figure 13, the two GBM models increase temperature estimation error over the random forest model, on average by less than 0.1°F for both models. Another important observation is that temperature estimation error is relatively stable below 2 meters, while typical infrared thermometers only support 10 to 50 centimeters [9].

6.2.2 Performance on Whole vs. Indoor Dataset. As recommended by ISO 13154:2017 [50], measurements are better standardized when people are acclimated to the indoor environment. To better study this phenomenon, we compare model performance when including or excluding people who are not acclimated to the indoor environment. During our deployment of *SIFTER* at the medical practice, the average outdoor temperature was 38°F compared to an average indoor temperature of 72°F, suggesting that people not acclimated to indoor temperatures may result in biased measurements.

As shown in Figure 14, each model's temperature MAE improves by 14.7%, 13.3%, 18.0%, 9.6%, and 14.7%, respectively, when using the indoor dataset vs. the whole dataset. As expected, the results indicate that the inclusion of measurements of people who are not acclimated to the indoor temperature reduces the ability for models to predict temperature. Baseline and average models achieve greater than 1 degree MAE on both datasets, and are not shown.

6.2.3 Improving Outdoor Acclimation Dataset Accuracy. One major challenge in automated fever screening systems is enforcement of acclimation to indoor environments. People who are screened before acclimation to the indoor environment will not be standardized with the indoor acclimated population, and as seen in Figure 14, results in higher errors between prediction and ground truth. Classification of acclimation can help in automated systems requiring acclimation to the indoor environment. Once detected, non-acclimated people can be flagged directly for a secondary screening, or made to wait for a certain time duration. We first explore improving the accuracy of temperature prediction by using ambient temperature and delay features. The intuition is that by improving temperature prediction, we can better differentiate whether people are acclimated to the indoor environment.

As shown in Figure 15, adding ambient temperature and delay features reduces temperature prediction error across all five regression models by 38.8%, 25.3%, 37.8%, 3.8%, and 1.6%. Random forest and gradient boosting methods benefit the most from the additional features. Using our best performing model, gradient boosting, we

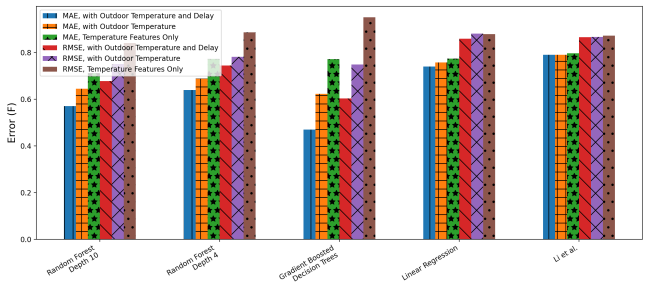


Figure 15: Mean Absolute Error of regression and baseline model temperature predictions using different feature sets.

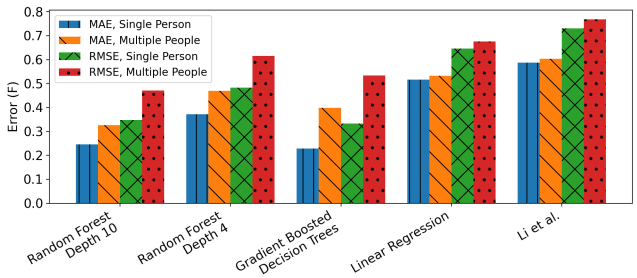


Figure 16: Mean Absolute Error of regression and baseline model temperature predictions on datasets of single person and multiple person measurements.

proceeded to apply a threshold to predict whether a person is acclimated with the addition of ambient temperature and delay features.

As shown in Table 2, the addition of outdoor ambient temperature provides an improvement in recall and F1 score for prediction of non-acclimation, while the addition of delay reduces both recall and F1 score. This suggests that while outdoor ambient temperature is useful for classifying acclimation, delay is only useful for predicting temperature. One explanation for this discrepancy is that delay does not have an effect on skin temperature of indoor acclimated people, but can help predict skin temperature changes as people acclimate to indoor environments.

Features	Precision	Recall	F1
No Extra Features	0.777	0.925	0.844
Outdoor Temperature	0.763	**0.984**	**0.859**
Outdoor Temperature, Delay	**0.847**	0.806	0.826

Table 2: Comparison of precision, recall, F1 score of the GBDT model on acclimation classification using different feature sets (default: distance, thermal features).

6.2.4 Performance with Multiple People. Another recommendation by ISO 13154:2017 [50] is to screen with only a single person at a time to minimize effects of multiple heat sources on the sensor's measurement accuracy. To better understand this effect, we compare the model performance with a single person or multiple people in the frame. We trained and evaluated the models on a 75% : 25% train test split on the whole dataset, and divided the evaluation results by number of people in frame. As shown in Figure 16, the MAE on the single dataset shows an improvement of 21.4%, 19.9%, 39.5%, 3.2%, and 3.3%, respectively, over the multiple dataset. However, with the additional error in multiple dataset, the overall error is still one order of magnitude less than the error of infrared-thermometer used in market which have an error of ±4°F [4]

6.3 Processing time

A low response time, i.e.,the time from the image captured by the RGB-thermal camera to the result shown on the client, is important to the real deployment. We measured the total image processing time for a single frame, which includes the entire pipeline mentioned in Section 3.

The average processing time in our local restaurant and hospital deployments are about 0.25 seconds and 0.231 seconds respectively, which is on par with a typical person's reaction time. The hospital deployment has a lower latency because the server it connects to utilizes a faster GPU. We note that a latency of even a few seconds is acceptable for our application because it typically takes a few seconds for a subject to walk across the observation window.

7 DISCUSSION

Deployment Considerations: NCITs, as well as IRT systems, are becoming more common in non-clinical settings to help combat the spread of COVID-19. However, they are not considered the gold standard for detecting fever in comparison with in-ear thermometers, oral thermometers, or axillary thermometers. Thus, one important factor is *how* to maximize the effectiveness of *SIFTER* in different environments while minimizing cost as an first-stage screening tool.

The primary consideration is the classical precision-recall tradeoff. Reducing false negatives allows for identifying a higher percentage of febrile humans, at the cost of falsely detecting more non-febrile humans. This would require more people to undergo a second-stage screening. On the other hand, reducing false positives reduces the number of people required to undergo a second-stage screening, at the expense of missing the detection of more febrile humans. Depending on the deployment, there may be a greater need for minimizing either false positives or false negatives.

SIFTER currently utilizes Light GBM with 5x asymmetric penalty. The threshold value that determines the boundary of febrile and non-febrile humans is set to first minimize false negatives, and second minimize false positives. This setting is more useful in low throughput settings such as small business and retail, where a first screen can reduce the number of people required to be screened with a secondary measurement. However, the threshold value can be raised for high throughput settings such as transportation hubs or commercial buildings to reduce the burden on administrators for a secondary screening. *SIFTER* allows the setting of this threshold value to allow the end users to choose an appropriate value for the deployment setting.

SIFTER requires a cloud server to process its pipeline, which leads to potential privacy concerns. *SIFTER* is intended for use at

[4] https://en.wikipedia.org/wiki/Infrared_thermometer

entrances and specific parts of buildings, rather than as a general-purpose surveillance camera throughout the building. For instance, the hospital where we conducted our study allowed us to deploy our system at its entrance with full knowledge of data locations. To preserve privacy, *SIFTER* does not store images on the server for more than 600 seconds and only saves bounding boxes and low resolution thermal images. In the future, we plan to improve privacy by moving major parts of our pipeline onto the edge, thereby removing the need to transmit raw data to a third-party server.

Procedure: As explored in Sections 6.2.2 and 6.2.4, there are four important factors that can affect temperature measurements: indoor environment acclimation, distance, presence of multiple people, and skin visibility. ISO 13154:2017 [50] recommends ensuring that people to be screened are acclimated to the indoor environment, are at a certain distance from the sensor, are screened one at a time, and relevant skin regions are visible to enable accurate measurements.

For automated fever screening systems, enforcement of these protocols is an important challenge. For indoor acclimation, we offer a method for helping to detect people who are not acclimated to the indoor environment in Section 6.2.3; however, this method is not applicable in all situations. For instance, people who have *mostly* acclimated to the indoor environment may not be distinguishable from fully acclimated people. In these situations, an extension to *SIFTER* that tracks a person's duration indoors to determine environment acclimation will provide additional robustness.

In Section 6.2.4, we noted that the temperature estimate errors in the presence of multiple people are not drastically different than for single person measurements, although there is an increase in error. One potential consequence of this study is a tradeoff between speed and accuracy. If multiple people can be screened at once, the throughput can be increased; however, there will likely be a sacrifice in accuracy. *SIFTER* provides a signal describing whether a single or multiple people are detected, which can be utilized by the client depending on the appropriate setting.

Distance is another critical parameter, and is usually enforced in modern IRT systems by placing a marker on the floor at the desired distance. As described in Section 6.2, distance has an effect on estimated temperature, although these errors can be mitigated by using distance as a feature to temperature models. This improvement in accuracy over distance reduces the need for occupants to directly position themselves for the sensor, thus reducing overall screen time and increasing throughput.

Finally, *SIFTER* and IRT systems in general require visibility of skin regions to measure temperature. In certain cases where relevant skin regions are not visible (such as due to facial coverings), we noted that this can be easily detected when facial region statistics fall below a definitive temperature threshold ($\sim 85^{\circ}$F). *SIFTER* also provides a signal noting these cases.

8 CONCLUSION

In this work, we demonstrate *SIFTER*, a low-cost system for continuously screening of people for fever without any human interaction. The thermal images are calibrated with a data-driven spot-size effect model to reduce the fundamental errors of thermal camera, and head orientation is estimated from RGB images with a retrained FSA-Net on a hand labeled restaurant dataset. The bounding boxes are matched between RGB and thermal images, and the calibrated thermal data is mapped onto a 3D point-cloud head model. We deployed *SIFTER* in two locations, a local restaurant and a medical practice. In our evaluations of this system on real patients in the medical practice, our system is able to achieve 100% TPR with only 22.5% FPR for screening occupants. This system can be used as an initial screening step (followed by additional screening of positive cases) to significantly reduce human labor cost of screening and social interaction.

ACKNOWLEDGMENTS

This research was partially funded by Columbia Engineering under the Technology Innovations for Urban Living in the Face of COVID-19 grant, and 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] Da Li, Carol C Menassa, and Vineet R Kamat. Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras. *Applied energy*, 251:113336, 2019.

[2] Rikke Gade and Thomas B Moeslund. Thermal cameras and applications: a survey. *Machine vision and applications*, 25(1):245–262, 2014.

[3] Yan Bai, Lingsheng Yao, Tao Wei, Fei Tian, Dong-Yan Jin, Lijuan Chen, and Meiyun Wang. Presumed asymptomatic carrier transmission of covid-19. *Jama*, 323(14):1406–1407, 2020.

[4] Esra Türe and Abdullah Yazar. How should we measure body temperature in the pediatric emergency department? which one is the most accurate? *Journal of Pediatric Infectious Diseases*, 14(03):121–126, 2019.

[5] Elena Chiappini, Sara Sollai, Riccardo Longhi, Liana Morandini, Anna Laghi, Catia Emilia Osio, Mario Persiani, Silvia Lonati, Raffaella Picchi, Francesca Bonsignori, et al. Performance of non-contact infrared thermometer for detecting febrile children in hospital and ambulatory settings. *Journal of clinical nursing*, 20(9-10):1311–1318, 2011.

[6] Sujit Shinde, Swapna Agarwal, Dibyanshu Jaiswal, Avik Ghose, Sanjay Kimbahune, and Pravin Pillai. Thermotrak: smartphone based real-time fever screening: demo abstract. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 639–640, 2020.

[7] Weijun Tan, Jingfeng Liu, Yue Zhuo, Qi Yao, Xing Chen, Wei Wang, Rushuai Liu, and Yu Fu. Fighting covid-19 with fever screening, face recognition and tracing. In *Journal of Physics: Conference Series*, volume 1634, page 012085. IOP Publishing, 2020.

[8] Armote Somboonkaew, Panintorn Prempree, Sirajit Vuttivong, Jutaphet Wetcharungsri, Supanit Porntheeraphat, Sataporn Chanhorm, Prasit Pongsoon, Ratthasart Amarit, Yuttana Intaravanne, Kosom Chaitavon, et al. Mobile-platform for automatic fever screening system based on infrared forehead temperature. In *2017 Opto-Electronics and Communications Conference (OECC) and Photonics Global Conference (PGC)*, pages 1–4. IEEE, 2017.

[9] Teledyne FLIR LLC. Technical note: Temperature guns versus thermal imaging technology. http://www.flirmedia.com/MMC/THG/Brochures/RND_048/RND_048_EN.pdf. [Online].

[10] Peter Wei, Yanchen Liu, Hengjiu Kang, Chenye Yang, and Xiaofan Jiang. A low-cost and scalable personalized thermal comfort estimation system in indoor environments. In *Proceedings of the First International Workshop on Cyber-Physical-Human System Design and Implementation*, CPHS21, page 1–6, New York, NY, USA, 2021. Association for Computing Machinery.

[11] Juhi Ranjan and James Scott. Thermalsense: determining dynamic thermal comfort preferences using thermographic imaging. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1212–1222, 2016.

[12] Peter Wei, Chenye Yang, and Xiaofan Jiang. *Low-Cost Multi-Person Continuous Skin Temperature Sensing System for Fever Detection: Poster Abstract*, page 705–706. Association for Computing Machinery, New York, NY, USA, 2020.

[13] Peter Wei, Yanchen Liu, and Xiaofan Jiang. Low-cost, perspective invariant and personalized thermal comfort estimation: Poster abstract. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks*

(Co-Located with CPS-IoT Week 2021), IPSN '21, page 396–397, New York, NY, USA, 2021. Association for Computing Machinery.

[14] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4), 2022.

[15] Stephen Xia, Jingping Nie, and Xiaofan Jiang. Csafe: An intelligent audio wearable platform for improving construction worker safety in urban environments. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, IPSN '21, page 207–221, New York, NY, USA, 2021. Association for Computing Machinery.

[16] Jingping Nie, Yanchen Liu, Yigong Hu, Yuanyuting Wang, Stephen Xia, Matthias Preindl, and Xiaofan Jiang. Spiders+: A light-weight, wireless, and low-cost glasses-based wearable platform for emotion sensing and bio-signal acquisition. *Pervasive and Mobile Computing*, 75:101424, 2021.

[17] Jingping Nie, Yigong Hu, Yuanyuting Wang, Stephen Xia, and Xiaofan Jiang. Spiders: Low-cost wireless glasses for continuous in-situ bio-signal acquisition and emotion recognition. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 27–39, 2020.

[18] Yigong Hu, Jingping Nie, Yuanyuting Wang, Stephen Xia, and Xiaofan Jiang. Demo abstract: Wireless glasses for non-contact facial expression monitoring. In *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 367–368, 2020.

[19] Stephen Xia, Daniel de Godoy Peixoto, Bashima Islam, Md Tamzeed Islam, Shahriar Nirjon, Peter R. Kinget, and Xiaofan Jiang. Improving pedestrian safety in cities using intelligent wearable systems. *IEEE Internet of Things Journal*, 6(5):7497–7514, 2019.

[20] Stephen Xia, Daniel de Godoy, Bashima Islam, Md Tamzeed Islam, Shahriar Nirjon, Peter R. Kinget, and Xiaofan Jiang. A smartphone-based system for improving pedestrian safety. In *2018 IEEE Vehicular Networking Conference (VNC)*, pages 1–2, 2018.

[21] Stephen Xia, Peter Wei, Jordan Misael Vega, and Xiaofan Jiang. Spindles+: An adaptive and personalized system for leg shake detection. *Smart Health*, 9-10:204–218, 2018. CHASE 2018 Special Issue.

[22] Daniel de Godoy, Stephen Xia, Wendy P. Fernandez, Xiaofan Jiang, and Peter R. Kinget. Demo abstract: An ultra-low-power custom integrated circuit based sound-source localization system. In *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 314–315, 2018.

[23] Daniel de Godoy, Bashima Islam, Stephen Xia, Md Tamzeed Islam, Rishikanth Chandrasekaran, Yen-Chun Chen, Shahriar Nirjon, Peter R. Kinget, and Xiaofan Jiang. Paws: A wearable acoustic system for pedestrian safety. In *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 237–248, 2018.

[24] Stephen Xia, Yan Lu, Peter Wei, and Xiaofan Jiang. Spindles: A smartphone platform for intelligent detection and notification of leg shaking. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, page 607–612, New York, NY, USA, 2017. Association for Computing Machinery.

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

[26] Ji Jia, Chengtian Xu, Shijia Pan, Stephen Xia, Peter Wei, Hae Young Noh, Pei Zhang, and Xiaofan Jiang. Moisture based perspiration level estimation. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '18, page 1301–1308, New York, NY, USA, 2018. Association for Computing Machinery.

[27] Dezhi Hong, Ben Zhang, Qiang Li, Shahriar Nirjon, Robert Dickerson, Guobin Shen, Xiaofan Jiang, and John A. Stankovic. Demo abstract: Septimu — continuous in-situ human wellness monitoring and feedback using sensors embedded in earphones. In *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 159–160, 2012.

[28] Stephen Xia, Rishikanth Chandrasekaran, Yanchen Liu, Chenye Yang, Tajana Simunic Rosing, and Xiaofan Jiang. A drone-based system for intelligent and autonomous homes. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, SenSys '21, page 349–350, New York, NY, USA, 2021. Association for Computing Machinery.

[29] Stephen Xia and Xiaofan Jiang. Improving acoustic detection and classification in mobile and embedded platforms: Poster abstract. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)*, IPSN '21, page 402–403, New York, NY, USA, 2021. Association for Computing Machinery.

[30] Stephen Xia and Xiaofan Jiang. Pams: Improving privacy in audio-based mobile systems. In *Proceedings of the 2nd International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things*, AIChallengeIoT '20, page 41–47, New York, NY, USA, 2020. Association for Computing Machinery.

[31] Peter Wei, Stephen Xia, Runfeng Chen, Jingyi Qian, Chong Li, and Xiaofan Jiang. A deep-reinforcement-learning-based recommender system for occupant-driven energy optimization in commercial buildings. *IEEE Internet of Things Journal*, 7(7):6402–6413, 2020.

[32] Peter Wei, Xiaoqi Chen, Jordan Vega, Stephen Xia, Rishikanth Chandrasekaran, and Xiaofan Jiang. A scalable system for apportionment and tracking of energy footprints in commercial buildings. *ACM Trans. Sen. Netw.*, 14(3–4), nov 2018.

[33] Ji Jia, Jinyang Yu, Raghavendra Sirigeri Hanumesh, Stephen Xia, Peter Wei, Hyunmi Choi, and Xiaofan Jiang. Intelligent and privacy-preserving medication adherence system. *Smart Health*, 9-10:250–264, 2018. CHASE 2018 Special Issue.

[34] Peter Wei, Stephen Xia, and Xiaofan Jiang. Energy saving recommendations and user location modeling in commercial buildings. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP '18, page 3–11, New York, NY, USA, 2018. Association for Computing Machinery.

[35] Peter Wei, Xiaoqi Chen, Jordan Vega, Stephen Xia, Rishikanth Chandrasekaran, and Xiaofan Jiang. Eprints: A real-time and scalable system for fair apportionment and tracking of personal energy footprints in commercial buildings. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, BuildSys '17, New York, NY, USA, 2017. Association for Computing Machinery.

[36] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5325–5334, 2015.

[37] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. From facial parts responses to face detection: A deep learning approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3676–3684, 2015.

[38] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[39] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[40] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5525–5533, 2016.

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

[42] Tuan-Hung Vu, Anton Osokin, and Ivan Laptev. Context-aware CNNs for person head detection. In *International Conference on Computer Vision (ICCV)*, 2015.

[43] Quang Huy Tran, Dongyeob Han, Choonghyun Kang, Achintya Haldar, and Jungwon Huh. Effects of ambient temperature and relative humidity on subsurface defect detection in concrete structures by active thermal imaging. *Sensors*, 17(8):1718, 2017.

[44] Yasir Dawood Salman, Ku Ruhana Ku-Mahamud, and Eiji Kamioka. Distance measurement for self-driving cars using stereo camera. In *International Conference on Computing and Informatics*, volume 1, pages 235–242, 2017.

[45] Jia Xiang and Gengming Zhu. Joint face detection and facial expression recognition with mtcnn. In *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, pages 424–427. IEEE, 2017.

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

[47] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 397–403, 2013.

[48] Dounia Bitar, Aicha Goubar, and Jean-Claude Desenclos. International travels and fever screening during epidemics: a literature review on the effectiveness and potential use of non-contact infrared thermometers. *Eurosurveillance*, 14(6):19115, 2009.

[49] Francis Ring and James Mercer. Thermal imaging for fever screening. *ISO Focus*, 1:33–35, 2007.

[50] ISO/TR 13154: 2009 ISO/TR 8-600. Medical electrical equipment–deployment, implementation and operational guidelines for identifying febrile humans using a screening thermograph. international standards organisation report, 2017.