Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A Novel Drone-based System for Accurate Human Temperature Measurement and Disease Symptoms Detection Using Thermography and AI

Ahmed Al Maashri¹, Ashraf Saleem²*, Hadj Bourdoucen¹, Omer Eldirdiry¹, Ahmed Al Ghadani³

¹Department of Electrical and Computer Engineering, College of Engineering, Sultan Qaboos University, Muscat, Oman
²Department of Applied Computing, College of Computing, Michigan Technological University, Michigan, USA
³Robotics and Artificial Intelligence Solutions Enterprise (RAISE), Muscat, Oman.

Abstract-

The world continues to witness several waves of COVID-19 spread due to the emergence of new variants of the SARS-CoV-2 virus. Stopping the spread requires synergistic efforts that include the use of technologies such as unmanned aerial vehicles and machine learning. This paper presents a novel system for detecting disease symptoms from a distance using unmanned aerial vehicles equipped with thermal and visual image sensors. A hardware/software system that uses thermography to accurately calculate the skin temperature of targeted individuals using thermal cameras is developed. In addition, machine vision algorithms are developed to recognize human actions such as coughing and sneezing, which are paramount symptoms of respiratory infections. The proposed system is implemented and tested in outdoor environments. The results of experiments showed that the system can determine the skin temperature of multiple targeted individuals simultaneously with an error of less than 1 °C. The field experiments showed that the developed system is capable of simultaneously measuring the temperature of more than 10 individuals in less than 5 seconds. Just to give a perspective, it takes at least 3 seconds to measure one individual’s temperature if this was done using traditional methods. Furthermore, the results showed that the system has accurately detected actions such as coughing and sneezing with almost 96% accuracy at a real-time performance of 28 frames/second.
1. Introduction

On January 30th, 2020, the World Health Organization (WHO) has declared COVID-19 a Public Health Emergency of International Concern (PHEIC) [1]. Not long after, WHO has declared that the disease has evolved into a global pandemic on March 11th, 2020 [2]. The disease has spread at a rate and scale rarely known to humanity. This pandemic is caused by a novel virus, officially named by the International Committee on Taxonomy of Viruses (ICTV) as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [3]. In response to the pandemic, governments around the world have implemented several drastic measures to contain the virus and flatten the curve [4]. Such measures include locking down cities, banning travel, and enforcing social distancing [5, 6, 7]. Since then, several variants of the virus have surfaced, including Delta and Omicron.

Individuals who get infected with SARS-CoV-2 show several symptoms that include fever, dry cough, tiredness, and difficulty in breathing [8]. These symptoms vary in occurrence and severity from one individual to another. One of the early studies published on COVID-19 by Huang, C. et.al [9] showed that infected patients exhibit certain symptoms with fever as the most prevalent feature in 98% of the patients, cough in 76%, and fatigue in 44%. It is evident that the early detection of these symptoms helps take appropriate actions in reducing the transmission rate of the disease.

Traditional methods used for detecting the symptoms of COVID-19 are time-consuming and costly, especially when conducted on a large number of individuals. In addition, there is a risk of spreading the infection to the medical staff as they get in close contact with potentially infected individuals. Furthermore, it is sometimes difficult to perform symptom inspections in densely-populated residential areas. Fortunately, technology can help in detecting and reporting suspected cases of SARS-CoV-2 infections. In fact, it was reported that the use of emerging technologies such as the Internet of Things (IoT), Unmanned Aerial Vehicles (UAVs), and Artificial Intelligence (AI) can provide fast and practical solutions to minimize the impact of the COVID-19 pandemic and leverage the burden on healthcare systems [10].

Furthermore, there is a need for detecting symptoms of COVID-19 in outdoor environments. An example of such environments is civil construction fields, where a large number of employees work close to each other and might transmit the virus in the working places. Such environments mandate the use of innovative solutions to remotely sense and detect the symptoms.

This paper describes a system that employs unmanned aerial vehicles (also known as drones) reinforced by remote sensing and AI to identify and report suspected cases of COVID-19 infections. This drone-based system is capable of analyzing a stream of visual and thermal images acquired from a low-accuracy thermal camera. The system measures the temperature and recognizes the actions of individuals from a distance in real-time.

However, several challenges are associated with the development of such systems. First of all, thermal image sensors used in outdoor environments have a large margin of error due to noise and reflections [11]. For example, the FLIR Zenmuse XT2 [12] has an accuracy of ±5 °C. This is way beyond the acceptable margin of error for skin temperature, which is ±0.5 °C. Another challenge
is that recognizing symptoms such as coughing and sneezing requires the ability to detect subtle gestures that happen in real-time. These challenges can be addressed using AI algorithms.

Figure 1 shows a pictorial representation of the developed system. A drone that is equipped with a dual thermal/visual camera is deployed to the targeted area. The drone pilot will point the camera to the concentrations of the people and capture images and videos. These images and videos will be transferred to the cloud, where several AI algorithms are used to process the stream of thermal and visual images. Some of these algorithms perform infrared thermography calculations on the thermal images. Other algorithms perform machine vision tasks such as action recognition and face detection. The system sends a notification when a suspected case is detected. The notification will include the temperature of the suspected case and the location. To the best of our knowledge at the time of writing this paper, there is no similar system that provides all the above-mentioned features that are available in our developed system.

The main objective of the research presented in this paper is to develop a hardware/software system that can detect suspected cases of SARS-CoV-2 infections, or any similar symptomatic infection disease, from a distance in real-time. The paper makes the following contributions:

- As mentioned earlier, one of the challenges is that outdoor thermal cameras have low accuracy. The paper employs the science of infrared thermography [13] to accurately
calculate the skin temperature of targeted individuals using thermal images. A novel self-calibration algorithm was developed to reduce the margin of error down to less than 1 °C.

- Accurate and real-time recognition of COVID-19 symptoms such as coughing and sneezing through the use of deep learning algorithms.
- Rapid and accurate detection of faces using machine vision algorithms. The challenge lies in the fact that individuals may wear masks that partially hide their faces.

The remaining of the paper is organized as follows: Section 2 presents the proposed system architecture for detecting COVID-19 symptoms using drones and infrared thermography. The novel self-calibration algorithm is described in Section 3, while Section 4 discusses the datasets acquisition and compilations processes. Experimental results are discussed in Section 5 and the paper is concluded in Section 6.

2. System architecture

The identification of possible COVID-19 infections is based on detecting the most prevalent features of infected patients as documented by Huang, C. et.al in their aforementioned research [9]. These symptoms include fever, coughing, nose-blowing, and sneezing. This section describes the developed system used to detect these symptoms. Figure 2 depicts a block diagram that describes the developed system. The system consists of three main subsystems:

1. **Skin Temperature Measurement**: this subsystem is responsible for measuring the forehead temperature of the targeted individuals from a distance.

2. **Human Action Recognition**: this subsystem recognizes actions such as nose blowing, sneezing, and coughing from a distance.

3. **Aggregator**: this subsystem aggregates the output of the two other subsystems. The aggregator consists of customizable high-level heuristics used for decision-making and reporting suspected cases. Additionally, the aggregator provides the visualization to the user.
Figure 2. The architecture of the developed system. The system consists of three subsystems; namely, Skin Temperature Measurement, Human Action Recognition and Aggregator.

2.1 Skin Temperature Measurement subsystem

Depending on the thermographic camera’s specifications, the range of the detected radiation in the infrared range can vary [14]. The produced images and videos from the thermal camera can be used for immediate diagnosis or processed through specialized software for further evaluation. Mounting the thermal camera in an unmanned vehicle is found to be a powerful combination for inspection in many applications, such as building monitoring, subsurface water leaks detection, oil and gas inspections, and search and rescue applications [15, 16]. However, most of the thermal cameras used with drones have a low-accuracy measurement, which does not meet the accuracy requirement (less than 1 °C) for skin temperature measurement. Therefore, a calibration algorithm is developed in this work that improves the accuracy to less than 1 °C.

Thermal images can be complemented with visual images of the same scene. A visual camera will provide the information needed to detect objects and regions such as faces. Once a face is detected, the forehead can be extracted as a region of interest (ROI). Furthermore, the thermal camera will provide the temperature for the desired ROI.

In this subsystem, geometric calibration is required to correct any distortion due to camera lenses. The calibration process adopted in this project is based on Zhang’s method [17]. After that, thermal and visual images are aligned. This alignment is needed since there is a 3-cm displacement in the x-axis between the visual and thermal sensors of the camera. Figure 3 illustrates the misalignment between the thermal and visual images.
Figure 3. The visual and thermal images are captured simultaneously from the same angle of incidence. The images are misaligned because of the distortion in the camera. The faces of the volunteers were intentionally blurred to avoid exposing their identity.

Face detection and forehead localization are performed in order to extract the temperature from the forehead of the targeted individual. To reduce the error margin of the thermal camera, an automatic self-calibration is performed, which will be discussed in detail in Section 3. Finally, the temperature of the targeted individual is measured and reported.

2.2 Human Action Recognition Subsystem

For the visual identification of COVID-19, three symptoms were considered in the Human Action Recognition (HAR); namely, coughing, sneezing, and nose blowing. In an ideal scenario, the frequency of these activities should be accounted for in combination with the thermal COVID-19 symptoms, which was discussed earlier. The aforementioned activities were recognized using three different Convolutional Neural Network (CNN) structures. Two of which are three-dimensional (3D) CNN, while the other is a two-dimensional (2D) CNN. The 3D CNN structures that were considered in this study are the Two-Stream Inflated 3D CNN [18] and the 3D ResNet-34 CNN [19]. Additionally, the YOLOv4-Tiny, which is a 2D CNN structure, was also considered.

HAR has been a challenging problem and has gained great interest in the field of healthcare. It is considered a time-series problem and many researchers rely on analyzing sensor data such as accelerometers [20]. However, the use of such accelerometers with a large group of people is not a practical solution. The alternative is to rely on visual analysis. Interestingly, a recent survey shows that only 9% of the existing literature in this field is devoted to vision-based HAR systems [21]. The challenges faced in this domain include the intraclass variation and interclass similarity, complex and various backgrounds under real-world settings, multi-subject interactions, group activities, and low-quality videos. In this study, the visual information is received as a feed from a camera attached to a drone.

Given that humans perform activities over a finite time, this means that a machine learning algorithm must observe the action as it develops over a sequence of frames. Therefore, the
information in the frames is both in the 2D space and in time. This turns HAR into a Spatio-temporal three-dimensional (3D) problem. However, 3D CNNs proved to be computationally demanding. The good news is that 2D CNN structures have shown promising results in analyzing HAR, albeit to some extent [22]. The limitation of 2D CNN is that it will attempt to identify the action without any knowledge of what happened before or after the moment the action was captured. In such cases, the HAR problem is attempted with limited contextual awareness. Since the temporal information is missing, the action sequence is unknown. Symptoms like coughing and sneezing are actions that get executed in a certain sequence and take time to perform. 3D CNN structures consider the entire action sequence to make a prediction. The following sequence of video frames in Figure 4. shows the action of sneezing. The sneezing action takes less than two seconds from start to end. Each frame will include activities such as raising hands, raising elbows, covering the nose, sneezing, lowering hands, and lowering elbows. Typically, the whole action is captured in 50 – 60 frames. 2D CNN structures will analyze each frame individually and will not retain any temporal knowledge and thus might not give accurate nor consistent predictions.

![Figure 4. Different frames for sneezing action](image)

To put things in perspective, Figure 5. shows a comparison between 3D and 2D convolution. The figure shows the complexity of the mathematical operations in the 3D CNN.
2.3 Aggregator Subsystem

The Aggregator subsystem receives two inputs:

- The temperature of the individuals from the “Skin Temperature Measurement” subsystem.
- The action of the individuals from the “Human Action Recognition” subsystem.

The Aggregator examines these two inputs and executes a set of heuristics for decision making. For example, assume that the measured temperature of an individual is 36 °C and no other symptoms were recognized by the HAR. In this case, the Aggregator can safely conclude that this individual is not showing any signs of infection.

On the other hand, let us assume that the measured temperature of the individual is 38 °C and the HAR shows that this individual is coughing. In such a scenario, the Aggregator will report that this individual is suspected. Consequently, the face of the individual and the GPS coordinates of the location are stored for identification purposes by the authority.

However, there might be border cases where it is difficult to conclude whether an individual is showing symptoms or not. Consider for example a case of an individual with a temperature of 37.3° C, while the HAR has reported that the individual could be sneezing with a confidence of 80%. In such a case, there are three alternatives for a decision:

1. Report case as NOT potentially infected.
2. Report case as potentially infected.
3. Report case as inconclusive.

It is advised to report this case as inconclusive and require the operator to investigate the case further. A manual examination in such cases will reduce false positives. However, it should be stated here that it is seldom when these border cases happen.

The Aggregator generates two outputs:
1. **Visualization**: This is a simple dashboard that shows the temperature of measured individuals and the current actions performed by them.

2. **Report of suspected cases**: For each case that is suspected to be infected with COVID-19, the Aggregator shall store an image that contains the face of the individual along with the GPS coordinates.

Since the Aggregator relies on a set of heuristics to make decisions, this makes the Aggregator a flexible tool that can be easily reconfigured based on the requirements. These requirements can be set by the user of the system.

Figure 6 shows the working process of the proposed system. After acquiring the visual (RGB) and thermal images from the drone camera, both images are aligned using geometric calibration as described earlier. The system detects faces and localizes the forehead of each detected face in the visual image. The forehead temperature is measured from the thermal image and the automatic self-calibration algorithm is applied to correct the measurement and reduce the effect of the background noise. Simultaneously, the system detects persons and recognizes the actions performed by the targeted individuals. Depending on the outcomes of the temperature measurement and human action recognition, the system will report the case as infected, not infected, or inconclusive.
Figure 6. Working process of the proposed system. $T_{ref}$ represents a reference threshold, where if the measured temperature is larger than $T_{ref}$, then this is considered as fever. $S_{COVID19}$ represents a set of symptoms such as coughing, sneezing, and nose blowing.

3. Automatic Self-Calibration Algorithm

The calibration of thermal radiation instruments can be accurately performed using known sources of spectral radiation. In general, two sources of calculated spectral radiation are used, either synchrotron sources or blackbody sources [23]. In practice, blackbody radiation offers a feasible and attractive solution for instrument calibration due to its simplicity and low cost [24]. However, blackbody sources are not possible to be used in measuring temperature with a non-stationary thermal camera similar to the one used in drones.

Therefore, a novel self-calibration algorithm for non-stationary thermal cameras is proposed. In the proposed method, a correction factor is automatically calculated as a function of the nominal human temperature $T_{nominal}$ (assumed 36.5° C), the average temperature of all objects in the image $T_{avg}$, and the precision of the measured temperature $T_{IPRCN}$. Assuming $n$ objects in the
thermal image, the temperature of each object $T_i$ is measured, a correction factor $\varphi_i$ is automatically estimated, and the calibrated temperature for each object $\hat{T}_i$ is then calculated using equations (1 – 5):

$$\hat{T}_i = T_i + \varphi_i$$

(1)

where,

$$\varphi = \overline{T} + \mu_i$$

(2)

$$\overline{T} = T_{nominal} - T_{avg}$$

(3)

$$\mu_i = (T_{IPRC} - 1) \cdot (\alpha \cdot T_{avg})^\beta$$

(4)

$$T_{IPRCN} = 1 - \left[ \frac{T_i - T_{avg}}{T_{avg}} \right]$$

(5)

The key element in calculating the correction factor is the constant $\mu$ shown in Equation (2). This constant represents the magnitude and the sign of the correction (i.e. whether the correction is positive or negative). The sign of the correction factor is determined using the precision of the corresponding measured temperature, while the magnitude is calculated using the average temperature of all objects.

$\alpha$ and $\beta$ in Equation (4) are tuning parameters that should be optimized during the calibration experiments. They are used to compensate for the change in the background of the objects, ambient temperature and humidity, drone angle of incidence, and illumination intensity within the objects’ surroundings. Once these parameters are set to the optimal values, it is found that the adoption of the correction factor should provide adequate temperature measurement accuracy. It is worth mentioning that $\alpha$ and $\beta$ were optimized manually in this project. However, optimization algorithms such as Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) can be employed to find the best values for these parameters.

It was noticed that the radiation from the background of the objects has a large effect on $T_i$ especially if the difference between $T_i$ and the background temperature $T_{iback}$ exceeds a certain threshold $THR$. The effect of this environmental noise can be compensated by using the following filter:

$$T_i = \begin{cases} \frac{\gamma (T_{iback} - T_i)}{T_i} & |T_i - T_{iback}| \geq THR \\ T_i & o.w. \end{cases}$$

(6)

This filter eliminates the environmental noise if the difference between the subject temperature and the background temperature exceeds a predefined threshold ($THR$). This filter is designed in a
way that the measured temperature will be reduced by $\gamma\%$ (typically $10-20\%$) of the difference between the measured temperature and the background temperature. In this study, $\gamma$ and THR are set to be $0.1(10\%)$ and $6\, ^\circ$C, respectively. These parameters were tuned based on the dataset collected for this study.

To analyze the accuracy of the developed algorithm, the following metrics are used:

- **Average of differences**: the reference measurement of the volunteer and the one recorded by the system are compared. Then, the average of the difference between reference and recorded is calculated using Equation (20):

$$D = \frac{\sum^n_{i=1}|T_{ref,i} - T_{measured,i}|}{n} \quad (7)$$

Where $n$ is the number of samples, $T_{ref}$ is the reference temperature (ground truth temperature), which was measured using an accurate thermometer before conducting the experiment, and $T_{measured}$ is the temperature measured by the algorithm. The unit is in degrees Celsius. The lower this metric is, the better the accuracy.

- **Precision**: this metric is a measure of how accurate is the algorithm in detecting true positive cases (i.e. infected cases), compared to the number of false-positive cases (i.e. the system reports a case with fever, while in fact, it is not). Precision is calculated as in Equation (8). The higher this metric is, the better the accuracy.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (8)$$

- **Recall**: this metric is a measure of how many true positive cases were detected from the total number of infected cases. The Recall is calculated as in Equation (9). The higher this metric is, the better the accuracy.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (9)$$

- **F-measure**: this metric combines both the Precision and Recall into a single figure. The F-measure is calculated as in Equation (10). The higher this metric is, the better the accuracy.

$$\text{F-measure} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

Equations (7) – (9) are used to calculate the precision, recall, and F-measure, respectively [21]:
4. Dataset Acquisition and Compilation

The developed system is capable of:

1. Measuring the temperature from a distance.

2. Recognize human actions that may indicate symptoms of COVID-19.

Therefore, a dataset of images and videos was acquired and compiled to validate the accuracy of the developed algorithms in performing the tasks mentioned above. More specifically, two datasets were compiled:

1. Temperature Measurement dataset: this dataset includes both visual and thermal images of targeted individuals. The images were captured at different altitudes and angles of incidence. Additionally, the images were captured at different times of the day and in different locations. These variations in altitude, angle of incidence, time of day, and location provide more rigorous testing and validation. It is worth mentioning that the minimum altitude of the drone was 5 meters. This is to avoid the effects of air turbulence on the measurements.

Figure 7 shows the drone’s hovering zone while capturing images. The drone was flying at altitudes of 5 – 10 meters, maintaining a Euclidean distance of 7 – 18 meters from the measured targeted individuals. The camera was pointed away from the sun to avoid direct high intensity of illumination, which may affect the measurements.

![Figure 7. Drone's flying region for capturing images and videos for the dataset](image)
Over 3,000 images and videos were compiled into this dataset, which included photos of hundreds of volunteers. Figure 8 shows samples of images from this dataset.

![Temperature Measurement Dataset Samples](image1)

Figure 8. Samples of images from the Temperature Measurement dataset. The top images are visual (encoded in RGB) and the ones at the bottom are the corresponding thermal images.

2. **Action Recognition Dataset**: this dataset includes visual images and videos of individuals performing activities such as coughing, sneezing, and nose blowing. For this dataset, the drone was in a stationary position, while the volunteers are passing by the field of view of the camera. As they pass by, the volunteers would intentionally perform actions such as coughing and sneezing. In this part of the data collection, the drone’s altitude was fixed at 5 meters. Samples of captured images from this data set are illustrated in Figure 9.

![Action Recognition Dataset Samples](image2)

Figure 9. Samples of images from the Action Recognition dataset
5. Results and Discussions

This section discusses the results of experiments conducted on the developed system. The section is split into three sub-sections, each of which discusses the results obtained from collected datasets for face detection and forehead localization, skin temperature measurement, and human action recognition.

5.1 Face Detection and Forehead Localization

The temperature should be measured from the foreheads of the targeted individuals. Therefore, the objective of this stage is to locate the forehead of the individuals in the image. However, to locate the forehead, the face should be detected first. The detection of the face has two advantages; one of which is to locate the forehead as mentioned above. Another advantage is to identify the face of the targeted individual if suspected of COVID-19 infection.

Traditionally, the Viola-Jones algorithm is used to detect faces [25]. While this algorithm is very popular and not compute-intensive, however, the algorithm suffers from high false positives. Instead, a classifier was trained using the datasets that were compiled in this project. The dataset includes the faces of many individuals, where some of them are wearing masks. Some individuals wore headwear and even spectacles.

Histogram of Oriented Gradients (HoG) [26] was used as an image descriptor. The feature vectors that are extracted from the HoG descriptor were used to train a Support Vector Machine (SVM) classifier. This SVM classifier acts as a face detector. Note that the training and the generation of the face detector were done offline. Figure 10 shows the face detector that was generated after training the SVM classifier with the dataset.

![Figure 10](image.png)

*Figure 10. A visualization of the face detector, which was generated after training the SVM classifier with the dataset.*

It was found that the precision of detection is 0.979, while the recall is 0.676. Furthermore, the F-measure was found to be 0.8.

When a visual image is received in this stage, the face detector is used to detect all the faces in the image. Once a face is detected, it is scanned to locate the forehead. This is achieved by using the face landmarks to locate the area just above the eyes, which mainly constitute the forehead of the
targeted individual. Now that we have detected the faces and located the corresponding foreheads, we can map the coordinates of faces and foreheads to the corresponding coordinates in the thermal image. This allows extracting the thermal measurement of each individual at the right location. See Figure 11 for a sample output.

![Figure 11. A sample of face detection (green box) and forehead localization (red box). The faces of the volunteers were intentionally blurred to avoid exposing their identity.](image)

### 5.2 Skin Temperature Measurement

Two sets of experiments were conducted for testing the proposed automatic self-calibration algorithm. The first experiment was conducted with 22 volunteers during the evening time (between 5 – 6 PM). The experiment was conducted at different locations (paved road, off-road, and nearby houses). Figure 12 shows samples of the captured images during the experiment.

The proposed self-calibration algorithm was applied to all thermal images taken at the location to measure the temperature of the sample. The measured thermal temperature is compared with the true temperature value of each subject to validate the effectiveness of the proposed algorithm. Figure 13 shows the error in temperature for each subject from randomly selected images. This error is the difference between the temperature indicated by the thermal camera and the measured temperature that was obtained using a medical thermometer for each subject. The measurement error is ranging from 2.5 to 6.2 °C. This large error is due to the low measurement accuracy of the thermal camera used with the drone.
Figure 12. Experiment 1 was conducted on the university’s campus. The faces of the volunteers were intentionally blurred to avoid exposing their identity.

Figure 13. Temperature measurement error in four randomly selected images of eight targeted individuals (T1 – T8) without using the proposed self-calibration algorithm. Notice that some of the targeted individuals may not show in all four images.
After applying the proposed self-calibration algorithm, the temperature measurement error has improved dramatically. As shown in Figure 14, the maximum measurement error has dropped by 74.2%. Also, the maximum average error has dropped from 5 °C to 0.7 °C, as shown in Table 1. This improvement in temperature measurement accuracy should allow using drones with low-accuracy thermal cameras in applications that require a low margin of error of temperature measurement.

![Figure 14. Temperature measurement error in each image after applying the proposed self-calibration algorithm.](image)

*Figure 14. Temperature measurement error in each image after applying the proposed self-calibration algorithm. The figure shows the reduction in measurement error for the same eight individuals in Figure 12 (T1 – T8). Notice that some of the targeted individuals may not show in all four images.*

*Table 1. Comparison of temperature measurement error before and after adopting the proposed self-calibration algorithm*

| Image Number | Average temperature measurement error | Average temperature measurement error after using the proposed algorithm | Improvement (%) |
|--------------|---------------------------------------|---------------------------------------------------------------------|-----------------|
| Image 1      | 3.300                                 | 0.544                                                              | 83.5            |
| Image 2      | 3.667                                 | 0.710                                                              | 80.6            |
| Image 3      | 3.850                                 | 0.264                                                              | 93.2            |
| Image 4      | 5.050                                 | 0.489                                                              | 90.3            |

To further validate the skin temperature measurement method, a second experiment was conducted with different subjects and different locations. The second experiment was conducted with 14 volunteers during the morning time (9:00 – 9:30 AM). The experiment was conducted at three zones (shade, sidewalk, asphalt). These variations allow testing the system at different levels of radiation noise. Notice that the volunteers are wearing masks in this experiment. Figure 15 shows samples from the experiment.
Figure 15. Experiment 2 was conducted at a different location with three zones.

Figure 16 shows the error in temperature measurements before and after using the self-calibration algorithm. Different images are selected from different zones with different subjects. It is evident that the proposed algorithm has improved the temperature measurement accuracy dramatically in all scenarios under different circumstances.

As mentioned in section 3, the algorithm performance is evaluated by applying the matrices in Equations 7-10 on all acquired images in both experiments. It is found that the average difference in the first and second experiments are 0.473°C and 0.538°C, respectively. Further examination of the results of the first experiment indicates that the Precision, Recall, and F-measure were found to be 0.750, 0.600, and 0.667, respectively. However, it was not possible to calculate the same metrics for the second experiment since there were no true positive cases. As for the execution time, on average, the processing of a set (i.e. a visual image and a thermal image) takes 4.72 seconds. This time includes the time to detect the faces, locate the foreheads, perform thermal self-calibration, measure the skin temperature, and store the output. This execution is reasonable, given that the execution of the algorithms was done on a commodity machine. The execution time can be improved significantly if a GPGPU workstation is used.
Figure 16. Temperature measurement error in each image at different zones before and after applying the proposed self-calibration algorithm.

Figure 17 shows samples of the algorithm output for measuring the temperature of multiple samples simultaneously.
5.3 Human Action Recognition

Three approaches were used to perform HAR. These are, the Two-Stream Inflated 3D CNN, 3D ResNet, and YOLOv4-Tiny. Since the latter approach had an advantage over the two other approaches, we shall briefly describe how YOLOv4-Tiny was trained.

YOLOv4-Tiny implements a smart approach for object detection that allows it to achieve real-time performance. The approach is based on the CSPDarknet53 backbone that runs a single-stage detector, which reduces the computations. Additionally, YOLOv4-Tiny inherently provides localization information. The architecture allows the model to identify the action and assigns it to an agent. This agent represents the individual performing that action. To detect an action without the temporal information, YOLOv4-Tiny will have to be trained to recognize the frames that indicate the possibility of action occurrence. It should be stated, however, that the absence of temporal information leads to low accuracy when predicting actions that are very similar to one another. The dataset used for training YOLOv4-Tiny was manually annotated using Microsoft Visual Object Tagging Tool (MS VOTT). The training used 64 batches and one subdivision. The learning rate is set as 0.00261 with 6000 as the maximum number of batches.

Figure 18 shows the recognition of human action of the three approaches. Notice how the algorithms have correctly predicted the actions. The results indicate that YOLOv4 has two key advantages over other 3D CNN structures. The first is that YOLO localizes the targeted individual in the image. The second is that YOLO provides real-time performance. The results show that the algorithm was able to maintain a frame rate of 28.4 fps, with a prediction accuracy of 95.87%.
Figure 18. Results of HAR: (a) Two-stream Inflated 3D CNN, (b) 3D ResNet, (c) YOLOv4-Tiny
6. Conclusions

In this paper, a system for detecting COVID-19 symptoms, from a distance was developed. The system utilizes technologies such as drones and AI to analyze both thermal and visual images and videos. The system is composed of three subsystems; namely, Skin Temperature Measurement, Human Action Recognition, and Aggregator.

The temperature measurement made by the system has successfully reached an error of less than 1°C. This was achieved by developing an automatic self-calibration algorithm that can be adapted for or any other infection disease with similar symptoms to COVID-19. In addition, the system is capable of detecting actions such as coughing and sneezing with an accuracy of 95.87%. This performance is maintained at 28.4 frames per second, which makes it favorable for real-time applications.

For future extension of this work, the authors would like to investigate on other infection diseases and include other symptoms such as fatigue and experiment with the aggregator in making more intelligent decisions for a wider range of applications.

Acknowledgment:

The authors would like to express their sincere gratitude to the Ministry of Higher Education, Research, and Innovation for funding this project (Research Grant RC/COVID-ENG/ECED/20/01). Also, the investigators would like to thank Sultan Qaboos University for providing a conducive academic environment to carry out the research. In addition, the research team highly appreciates the support that the Ministry of Health and the Telecommunications Regulatory Authority have provided in the early stages of the research. Furthermore, the authors thank Data2Cloud for providing Cloud services for data storage and algorithm processing. Finally, the investigators appreciate the assistance provided by Frontier Tech LLC and its CEO, Mr. Yaqoob Al Hatali for conducting some of the field tests on and off-campus.

References

[1] WHO, "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)," 30 January 2020. [Online]. Available: https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov). [Accessed 9 10 2020].

[2] WHO, "WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020," 11 March 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. [Accessed 9 10 2020].
[3] M. Bchetnia, C. Girard, C. Duchaine and C. Laprise, "The outbreak of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): A review of the current global status," *Journal of Infection and Public Health*, pp. 1417-1427, 2020.

[4] WHO, "WHO Director-General's opening remarks at the media briefing on COVID-19 - 18 March 2020," 18 March 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---18-march-2020. [Accessed 11 April 2020].

[5] A. B. a. A. Loeb, "Flattening the COVID-19 Curves: Social distancing imposes hardships, but it can save many millions of lives," 17 March 2020. [Online]. Available: https://blogs.scientificamerican.com/observations/flattening-the-covid-19-curves/. [Accessed April 2020].

[6] A. Younes, "How Jordan is flattening its COVID-19 curve," 22 April 2020. [Online]. Available: https://www.aljazeera.com/news/2020/04/jordan-flattening-covid-19-curve-200422112212466.html. [Accessed April 2020].

[7] Oman Air, "Travel Advisory on COVID-19 coronavirus outbreak," [Online]. Available: https://www.omanair.com/om/en/travel-advisory-covid-19-coronavirus-outbreak. [Accessed April 2020].

[8] WHO, "Q&A on coronaviruses (COVID-19)," 17 April 2020. [Online]. Available: https://www.who.int/news-room/q-a-detail/q-a-coronaviruses. [Accessed April 2020].

[9] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang and B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497-506, 2020.

[10] V. Chamola, V. Hassija, V. Gupta and M. Guizani, "A Comprehensive Review of the COVID-19 Pandemic and the Role of IoT, Drones, AI, Blockchain, and 5G in Managing its Impact," *IEEE Access*, vol. 8, pp. 90225-90265, 2020.

[11] Y. Zefri, I. Sebari, H. Hajji and G. Aniba, "In-depth investigation of applied digital photogrammetry to imagery-based RGB and thermal infrared aerial inspection of large-scale photovoltaic installations," *Remote Sensing Applications: Society and Environment*, vol. 23, no. 100576, 2021.

[12] FLIR, "Zenmuse XT2 User Manual," 2018. [Online]. Available: https://www.flir.com/globalassets/imported-assets/document/zenmuse-xt2-user-manual.pdf. [Accessed April 2020].

[13] W. Minkina and S. Dudzik, *Infrared Thermography: Errors and Uncertainties*, Wiley, 2009.
[14] W. K. Wong, P. N. Tan, C. K. Loo and W. S. Lim, "An effective surveillance system using thermal camera," in *International Conference on Signal Acquisition and Processing*, Kuala Lumpur, Malaysia, 2009.

[15] S. O. Johnsen, T. Bakken, A. A. Transeth, S. Holmström, M. Merz, E. I. Grøtli, S. R. Jacobsen and R. Storvold, "Safety and security of drones in the oil and gas industry,", in *The 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*, Venezia, 2020.

[16] Y. Naidoo, R. Stopforth and G. Bright, "Development of an UAV for search & rescue applications," in *IEEE Africon’11*, Livingstone, Zambia, 2011.

[17] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999.

[18] J. Carreira and A. Zisserman†, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," February 2018. [Online]. Available: http://arxiv.org/abs/1705.07750. [Accessed 11 October 2020].

[19] K. Hara, H. Kataoka and Y. Satoh, "Learning Spatio-Temporal Features with 3D Residual Networks," August 2017. [Online]. Available: http://arxiv.org/abs/1708.07632. [Accessed 11 October 2020].

[20] J. Wang, Y. Chen, S. Hao, X. Peng and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3-11, 2019.

[21] D. R. Beddiar, B. Nini, M. Sabokrou and A. Hadid, "Vision-based human activity recognition: a survey," *Multimedia Tools and Applications*, 2020.

[22] S. Shinde, A. Kothari and V. Gupta, "YOLO based Human Action Recognition and Localization," *Procedia Computer Science*, vol. 133, pp. 831-838, 2018.

[23] J. Hollandt, J. Seidel, R. Klein, G. Ulm, A. Migdall and M. Ware, "Primary Sources for Use in Radiometry," *Optical Radiometry, Experimental Methods in the Physical Sciences*, vol. 41, no. 1, 2005.

[24] J. Hartmann, "High-Temperature Measurement Techniques for the Application in Photometry, Radiometry and Thermometry," *Physics Reports*, vol. 469, no. 5-6, pp. 205-269, 2009.

[25] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, HI, USA,, 2001.
[26] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005.
**Ahmed Al Maashri:** Conceptualization, Formal analysis, Validation, Investigation, Visualization, Project administration, Funding acquisition, Writing - Original Draft

**Ashraf Saleem:** Conceptualization, Data Curation, Formal analysis, Validation, Investigation, Funding acquisition, Writing - Original Draft

**Hadj Bourdoucen:** Methodology, Data Curation, Validation, Supervision, Funding acquisition, Writing - Review & Editing.

**Omer Eldirdiry:** Validation, Visualization, Software.

**Ahmed Al Ghadani:** Visualization, Software.
Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ahmed Al Maashri reports financial support was provided by Ministry of Higher Education, Research, and Innovation.