Advances in Consumer Research
https://acr-journal.com/
Volume-3 | Issue-7 |July: 2026

Original Researcher Article

Design and Performance Optimization of a Low-Cost IoT-Based Thermal Imaging System Using MLX90640 and ESP32

Dr. Rohit Sharma[1], Mr. Samarjeet Singh[2], Dr. Vibhash Yadav[3], Prof. Anita Yadav[4], Prof. Raghuraj Singh[5]
[1] Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology for Divyangjan, Uttar Pradesh, India.
[2] Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India.
[3] Department of Information Technology, Rajkiya Engineering College, Banda, Uttar Pradesh, India.
[4] Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India.
[5] Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India.
.

ABSTRACT
The benefits of cost-effective infrared imaging include the ability to measure temperature changes in a patterned way without using high-temperature resolution bolometers. In addition, it allows for remote measurement and processing. However, the full benefit of infrared imaging will not occur until after many years. Therefore, a low-cost, single-board IoT thermal imaging prototype was designed as a first step toward lower-cost, less-complex infrared sensors. The thermal image sensor selected was the 32 x 24 pixel far-infrared array MLX90640. It sends its measurements via I2C to a small microcontroller (ESP32) that then computes and displays the results either on a screen connected directly to the microcontroller or on a web page that uses HTTP requests to get the data.

In this article, four development versions of the sensor are described. Version one used direct color mapping of the temperature values. Version two added temporal exponential smoothing. Version three smoothed both temporally and spatially using a 3x3 filter. Finally, Version four included all features previously mentioned in addition to a minimum span scaler that normalized each frame to have a total span of 100°C, extracted hotspots based upon the standard deviation of temperature across the entire frame, and provided a choice of whether to send the data to a web server remotely, or if the user would like to save the data to a file locally.

Version four demonstrated a minimum temperature value of 30.07 °C, maximum temperature value of 35.77 °C, and temperature value of 34.99 °C which clearly indicated where the hotspot existed. These numbers also demonstrate a 5.7 °C frame span, 4.92 °C hotspot contrast relative to the minimum, and a normalized hotspot value of 0.863. The author's subjective estimate of how stable the system would operate increased from 3 out of ten in Version one to 9 out of ten in Version four. The analytical portion of the article explores the trade-offs that allowed for such improvements. Specifically, the authors discuss how reducing the weighting applied to older time stamps (using a 0.75 weighting) reduces noise by applying a Temporal Smoothing Filter with an ideal RMS reduction value of 0.378 compared to 0.333 if only a simple averaging filter were applied over 3x3 pixels. If the noise were truly independent, then both filters could be combined to achieve an even better noise-reduction factor of 0.126. Additionally, as the size of the averaging filter increases so too does the number of frames required to settle back to within 10 % of a step, providing a demonstrable stability response tradeoff....

Keywords: thermal imaging; MLX90640; ESP32; IoT; edge processing; temporal smoothing; spatial filtering; hotspot detection; TFT visualisation.

INTRODUCTION

Thermal imaging converts spatial differences in IR radiation into a temperature field that may be visually interpreted as an image. Unlike a visible light camera, thermal imaging may detect warm objects, people, and temperature gradients in complete darkness. These characteristics indicate that thermal imaging has applications for predictive maintenance, inspecting buildings, contactless screening, analyzing occupancy, fire prevention research, and private/anonymous activity monitoring. In addition to low cost, low resolution thermopile sensors are attractive for use in IoT nodes since they provide less visual identity than typical cameras yet retain the thermal shape necessary for detecting many types of anomalies [4],[5] (the "thermal signature").

While sensor price is often cited as a major barrier to implementation of thermal imaging systems in practice, there are other barriers including repeated measurement requirements, calibration requirements, the need for real time data filtering, a stable color representation, rapid feedback via display, and the requirement for a power supply that can accompany the device into the field. Commercial hand held cameras provide all of the above mentioned features but typically contain higher resolution sensors and proprietary processing algorithms. Designing a microcontroller based solution will reduce the overall system complexity but due to their limited memory and computational resources, high performance super-resolution techniques and/or high performance deep learning techniques cannot run at the sensor frame rate.

Therefore this project will evaluate how much usable thermal image quality can be generated by using very light weight signal processing methods versus expensive hardware. The MLX90640 offers a precalibrated 32x24 pixel sensor with adjustable frame rates and a standard I2C interface [1]. The ESP32 module offers Wi-Fi capabilities along with all the peripheral interfaces needed to read the sensor array and control a TFT display [2]. The ILI9341 offers a 240x320 color TFT display and supports both 8- and 16-bit pixel formats. Therefore together these modules offer a compact edge node capable of computing temperature fields from raw sensor readings. Signal processing occurs entirely within the ESP32 as opposed to a central server. The primary contribution of this study is the systematic engineering evaluation of an optimization path consisting of four different versions. While previous studies have demonstrated a single thermal image created by applying certain signal processing techniques to the output of a thermal sensor array [6],[7],[10],[11][14], this report evaluates which specific stages of signal processing produce measurable improvements in image quality, at what computational expense, and under what conditions such signal processing may lead to misleading interpretations. Furthermore, this report converts previously reported numerical parameter values used to create images presented in the authors' original manuscript into actual numerical values allowing reproduction. Since attractive colors do not prove correctness, prior reports must demonstrate: measured evidence from the prototype being tested, qualitative observations made during testing, and analytical predictions.

The particular contributions made in this study include;

A comprehensive but low-cost system design for the capture of data from the MLX90640 sensor using an ESP32 microcontroller, to render the captured temperature data on a TFT screen, and to monitor the collected data via a web-browser.

A staged methodological approach to process the acquired thermal data, which includes (i) raw mapping, (ii) exponential temporal smoothing, (iii) 3 x 3 spatial filtering, (iv) minimum-span adaptive scaling, (v) RGB565 conversion and (vi) hotspot detection.

A numerical evaluation of (i) frame buffer size, (ii) I2C data transfer time, (iii) filter settling time, and (iv) theoretical noise suppression.

A safe-results section with evidence as presented by the authors including all their published images and reported values. This is clear about no raw frames of temperature data being available nor instrument calibrated or reference measured.

2. Literature Review and Research Gap

2.1 Low-resolution thermopile sensing

The MLX90640 has a total of 768 far infrared sensors that are arranged in a grid of 32 columns with 24 rows. The manufacturer states that there will be factory calibrated data from the device, and that it will operate on a 3.3V power source. The manufacturer also states that the refresh rate of the device can be programmed, and that the noise equivalent temperature difference (NETD) of the device is approximately 0.1K rms @ 1Hz as long as specified conditions are met [1]. The NETD is a value that reflects how much error exists within each individual pixel, and thus represents a minimum level of accuracy per pixel. However, this does not guarantee overall system accuracy. After the sensor is installed into an enclosure and directed towards a true target object, the accuracy of the determined temperature is dependent upon many factors including: Emissivity of the target object; Reflected Apparent Temperature (i.e. objects behind or near the subject); Distance between the sensor and the subject; Atmosphere present around the sensor/subject pair; Pixel Fill Factor of the sensor element; Optical Field of View (FOV) of the sensor element; Self Heating Effect due to electrical operation of the sensor; And the numerical reconstruction process applied to the raw data by the host computer.

Riquelme et al. developed a privacy compliant fall detection using the same 32 x 24 MLX90640. Riquelme found that even though personally identifiable visual information was removed from the images, coarse posture could still be identified [4]. Rodríguez-Cobo et al. compared multiple thermopile arrays and chose the MLX90640 for use in a non-contact diagnostic application based on the pixel density and stability of the device over other options evaluated; their method of calibrating the device to two points demonstrates that a reference temperature is necessary if accurate measurement of absolute temperatures is desired [5]. Alves et al. demonstrated how a low cost thermal camera could aid in reducing the computational overhead associated with determining pose by providing a small thermal cue early in the processing pipeline [6]. Collectively, these studies provide justification for the selection of this particular sensor type. They also demonstrate that output from this sensor must be processed as a measured field which requires calibration and processing prior to being used as input to subsequent image analysis algorithms, rather than being treated simply as another form of photography.

2.2 Privacy-preserving monitoring and activity interpretation

A number of researchers have explored the use of low resolution infrared cameras for monitoring people without disturbing them. Tateno et al. (2017) integrated a 2D Infrared camera using 3-Dimensional Convolutional Network (CNN) for Fall Detection. Yin et al. (2018) employed Recurrent Neural Networks to preserve the Temporal Structure of Thermal Sequences from Low Resolution Cameras. Márquez et al. (2000), researched Activity and Fall Classification in Older Adults Using Non-Invasive Infrared Camera Sensors. Muthukumar et al. (2015) investigated Denoising and Super Resolution Prior to Activity Recognition, specifically, identified Blur, Noise and Environmental Transfer as Realistic Limitations of Low Resolution Infrared Systems.

2.3 Edge processing, occupancy and system integration

Perra et al. demonstrated in their paper that doorway occupancy could be monitored using low-cost infrared array sensors by processing the data on site, which indicated that effective decisions can be made when not sending traditional video over networks [14]. Fraga-Lamas et al. showed that placing thermal sensor technology in a hybrid (edge/MIST) system will allow the cost, resiliency of the network, amount of communication required for each node and computation at the node to be balanced as part of an overall solution for industrial safety [15]. Djaja-Josko et al. improved the accuracy of indoor localization through the use of fusion of far-infrared arrays with proximity sensors [8]. Perov and Heger provided evidence that the choice of interpolation method and kernel used in convolution affects downstream localisation performance [16], supporting the use of edge computing. However, the authors also point out that display enlargement, filtering, and classification represent different functions; while interpolation may provide continuity for classifiers or visualization, adding no new sample points from the physical detectors.

2.4 Image enhancement and embedded constraints

The cited literature, from which most of the original content was derived, includes temporal smoothing, spatial filtering, interpolation, RGB565 rendering, hotspot localization and browser visualization as the key optimization areas. However, citation verification revealed that many of the provided DOIs and titles were paired to articles discussing completely different subjects. Therefore, in order to maintain academic accountability for the material presented here (i.e., the technology used), the following review will retain the originally intended technical categories, while restricting their citations only to those articles with both a verifiable bibliographic reference and subject matter. Similarly, the second paper on face-anti spoofing is also treated separately, such that only the authors' ablation methodology (component-wise) is adapted into the V1-V4 experimental structure [7] for thermal imaging analysis. Daim et al. developed a valid bibliometric method for predicting new technologies [17]; although this represents a viable method for horizon scanning it does not relate directly to evaluating temperature, latency or image quality. Thus this area is omitted within the performance rationale.

Ultimately, for an ESP32-based system design, the critical factor determining whether a particular enhancement technique may be applied effectively is whether its associated memory, arithmetic, timing and error characteristics allow for predictable operation within a fixed frame loop. Deep super resolution techniques have value in enhancing images when sufficient data exist in a training set and a suitable computing platform is available [12], while a simple one-pass (first order) temporal smoothing filter and a 3x3 neighborhood filter are relatively low complexity and do not depend upon large amounts of data or complex processing resources. Therefore the current research focuses on identifying methods which operate in a transparent manner and have effects that can be determined through analytical means and thus can be inspected at the pixel level.

2.5 Comparative synthesis of verified literature

Ref.	Study / year	Focus and method	Principal contribution	Limitation relevant to this work
[4]	Riquelme et al., 2019	MLX90640 fall dataset	Demonstrated privacy-friendly 32 x 24 thermal sequences	Dataset task; not a complete ESP32 display study
[5]	Rodríguez-Cobo et al., 2024	Non-contact sensing; sensor comparison and two-point calibration	Selected MLX90640 and demonstrated the value of in-range calibration	Final AI workload migrated to a Raspberry Pi
[6]	Alves et al., 2023	Thermal-assisted pose detection	Used a low-cost array to reduce CPU workload	Application-specific; limited display-pipeline detail
[8]	Djaja-Josko et al., 2025	Thermal/proximity sensor fusion	Improved localisation by adding FIR-array evidence	Multi-sensor localisation rather than thermography

Ref.	Study / year	Focus and method	Principal contribution	Limitation relevant to this work
[9]	Tateno et al., 2020	3D-CNN fall recognition	Showed privacy-preserving sequence classification	Learning and deployment complexity exceeds the present filter chain
[10]	Yin et al., 2021	LSTM activity recognition	Modelled temporal information in low-resolution IR data	Requires labelled sequences and trained inference
[11]	Márquez et al., 2022	Older-adult activity/fall classification	Demonstrated non-invasive monitoring potential	Application cohort and sensor placement constrain generalisation
[12]	Muthukumar et al., 2022	Denoising and super-resolution before DL	Connected image enhancement with recognition quality	Higher computational and training cost
[13]	Newaz & Hanada, 2024	Privacy-preserving activity recognition	Retained activity cues without visible-camera identity	Recognition study; no calibrated thermography claim
[14]	Perra et al., 2021	Doorway occupancy with local processing	Demonstrated real-time edge decisions and private sensing	Constrained geometry and counting task
[15]	Fraga-Lamas et al., 2022	Industrial edge/mist thermal safety	Integrated thermal sensing with resilient IoT architecture	Uses a higher-resolution thermal core for the main system
[16]	Perov & Heger, 2024	Interpolation and CNN localisation	Quantified the effect of preprocessing on downstream decisions	Classifier gains do not imply higher thermal measurement resolution

In addition to the findings from the synthesis, there are three additional consistent differences in what was found when studying applications. The first difference is that application accuracy is typically evaluated with respect to the output of some sort of visualization process, but the visualization process itself is rarely documented completely. The second difference is that if visualizations are improved either as far as how well they appear or as far as their ability to classify objects within them (i.e., by using techniques like super resolution), then it may be at the expense of blurring peak values in images and/or moving boundaries further away than expected. The third difference is that most prototype systems that have been studied, did not explicitly distinguish between those properties of sensors which were determined through specification (e.g., sampling rate) versus measurements of the actual system being tested versus predictions derived from models.

2.6 Research question and design objective

The primary research question is how well does simple, deterministic (low complexity) signal processing assist in improving both the consistency and readability of an ESP32 MLX90640 generated heatmap?

In other words, as much as possible, what will be the effect on spatial resolution, and metrologically accurate measurements when this processing methodology is utilized?

An associated goal for this project would be to create a mobile acquisition-to-visualization path, to determine the theory-based response and resource cost of that path, and then to determine which measurements are needed for a valid verification.

This position is narrower than a medical diagnostic study, or one that focuses on developing some sort of deep learning based solution; however, the work being done here has direct relevance to researchers that have to make a constrained thermal node behave in a predictable way prior to adding a specific classifier for their task.3. Proposed System Architecture

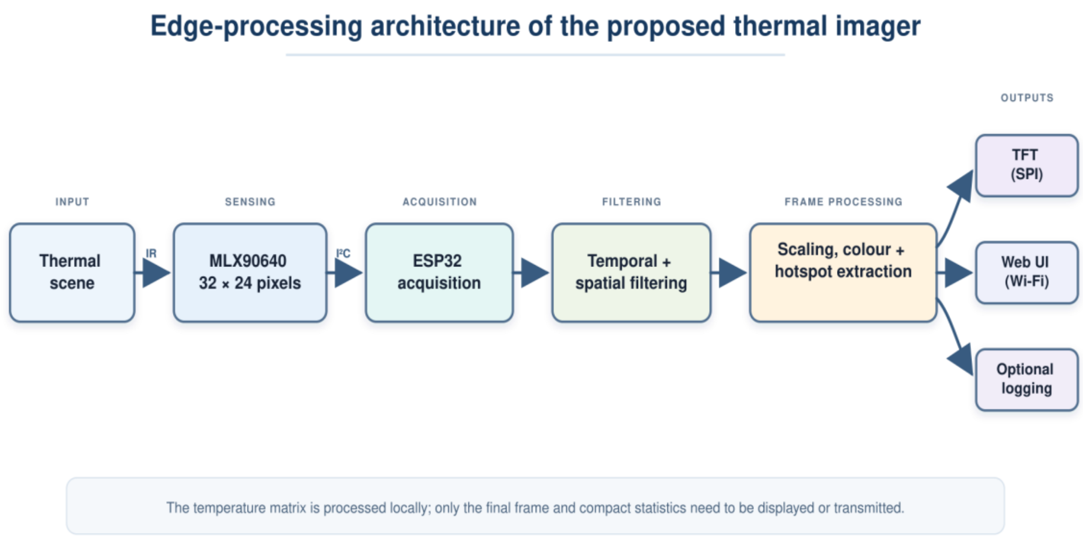

Figure 1. Edge-processing architecture of the proposed MLX90640-ESP32 thermal imaging system.

3.1 Hardware selection

Element	Role in the prototype	Design reason / constraint
MLX90640	Produces a 32 x 24 temperature matrix (768 values)	Low-resolution, factory-calibrated FIR array with I2C interface; suitable for compact nodes
ESP32	Acquisition, calibration calculations, filtering, hotspot extraction, Wi-Fi and control	Integrated connectivity and sufficient SRAM/compute for lightweight per-pixel operations
ILI9341 TFT	Local 240 x 320 colour heat-map display	SPI-compatible controller; RGB565 reduces display traffic to 16 bits per rendered pixel
2,000 mAh Li-ion battery	Portable energy source	Allows stand-alone use; actual runtime depends on display brightness, Wi-Fi traffic and conversion losses
Web browser	Remote view and thermal statistics	Avoids a dedicated client application; any device on the permitted network can display the interface

The MLX90640 is accessed by the ESP32 via an I2C interface (the frequency of which was set to 400 kHz for the purpose of this project), as is the case with the TFT. The reason for using SPI with the TFT is that it produces significantly more data than required for reading sensor configurations. It also provides isolation between the sensor readout bus and screen rendering process. The WiFi connectivity will be limited to an App level (i.e. no

WiFi = no app connectivity) therefore losing WiFi access does not affect the functionality of local display.

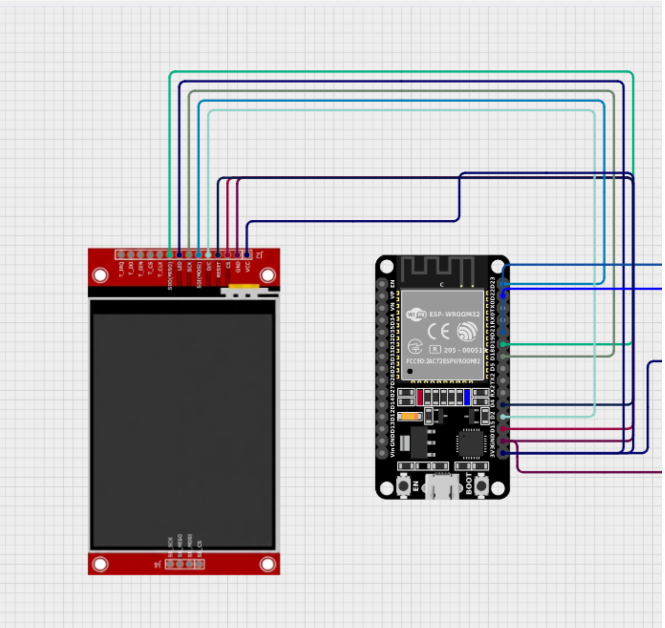

Figure 2. Author-supplied prototype wiring diagram showing the MLX90640, ESP32 and TFT interconnection. Pin assignments should be checked against the exact development-board revision before construction.

3.2 IoT layer model

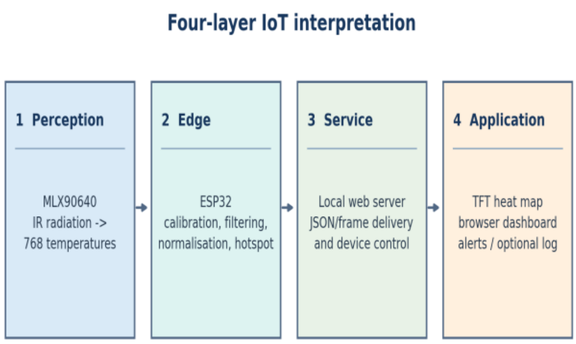

Figure 3. Four-layer interpretation of the proposed IoT system.

The "perception" or "sensing" layer converts IR energy to a numerical value using an Infrared Sensor at the sensor level. All of the latency sensitive functions occur in the "edge" or processing layer (i.e., reconstructing the temperature image, validating it, removing noise, converting temperature units and extracting hotspots). The "service" layer serves up data from this process as small HTML files via the ESP32 Web Server. The "application" layer contains everything else that would be part of your project; including the display for showing the heat map on the TFT screen, the browser based dashboard, alert messaging system, etc. By separating out the edge functionally from remote services you will reduce network delay time, and allow your project to still have some functionality even when no WiFi is present.

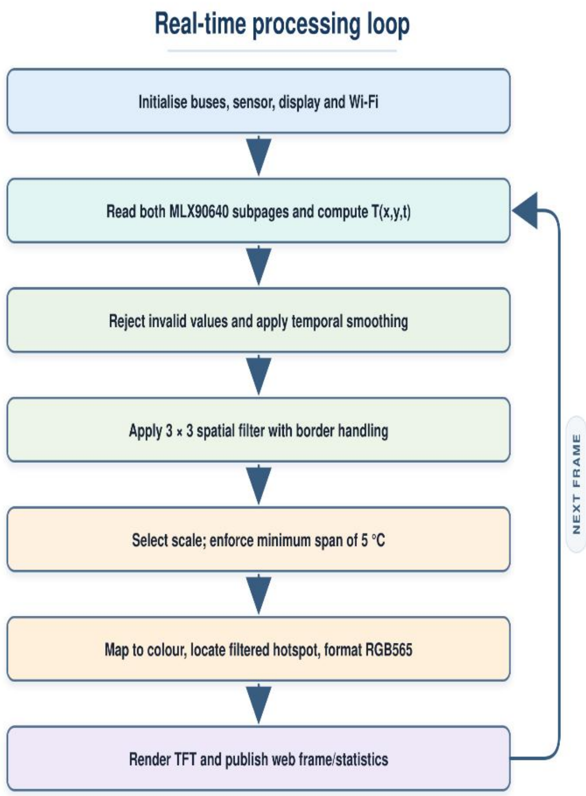

Figure 4. Processing sequence executed for each thermal frame.

4. METHODOLOGY

4.1 Infrared measurement model and frame acquisition

The surface at the temperature T is emitting radiant energy with an amount that is roughly related to the fourth power of temperature. The radiant energy portion for the idealized grey surface may be described by the Stefan Boltzmann relationship.

$$E = \varepsilon \sigma T^4$$

Where E = Radiant Exitance, ε = Emissivity, and σ = The Stefan Boltzmann Constant. The formula illustrates how radiant energy relates monotonically to absolute temperature, however a compact radiative transfer model will represent what an array pixel would observe as a combination of radiant energies versus pure black body radiation.

$$L_{\mathrm{meas}} = \tau[\varepsilon L_{\mathrm{bb}}(T_{\mathrm{obj}}) + (1 - \varepsilon)L_{\mathrm{bb}}(T_{\mathrm{refl}})] + (1 - \tau)L_{\mathrm{bb}}(T_{\mathrm{atm}}) \qquad (1a)$$

The above equation represents how much of the emitted radiation reaches the sensor and also includes the component related to reflection at the surface. Where, L_{meas} = Radiance measured by the detector. L_{bb} = Blackbody radiance integrated over the detector's wavelength band. T_{obj}= Object temperature. T_{atm} =

Atmospheric temperature. T_{refl} = Apparent temperature of the reflecting surface. This last parameter is very important to include since it provides an effective way to measure shiny surfaces such as metallic, glossy plastics and glass. However, if too high an emissivity factor is provided for a low-e material, it will attribute the radiant energy reflected back from the surface to the target and result in an incorrect recovery of the actual temperature of the target. Hence, visual color contrast can still occur when there has been no correct recovery of the true temperature.

Since the sensor does not have a direct measurement capability using Equation (1a), it collects pixel data along with die-temperature and compensation-pixel data. Using stored calibration factors contained within EEPROM, it uses these calibration factors to calculate an estimated temperature for each pixel. Manufacturer supplied software, specifically a calculation procedure and a driver library should be referenced for this purpose [1].

The resulting frame at time t is represented by

$$T_t = \{T(x, y, t) | x = 1, \ldots, 32; y = 1, \ldots, 24\} \tag{2}$$

which contains N = 32 x 24 = 768 floating-point temperature values. With 32-bit floats, one temperature frame occupies 768 x 4 = 3,072 bytes. Retaining the current and previous filtered frame requires about 6,144 bytes before display or network buffers are counted.

4.2 Version 1 - direct thermal mapping

Version 1 maps each raw temperature estimate directly to a normalised intensity. For a frame with extrema T_{min} and T_{max}

$$I(x, y) = clip\left[\frac{T(x, y) - T_{min}}{T_{max} - T_{min}}, 0, 1\right]$$

and $I(x, y)$ is passed to a colour lookup table. This method has O(N) complexity, needs no history buffer and responds immediately to temperature changes. Its weakness is equally direct: any frame-to-frame perturbation in a pixel or in the two extrema changes the displayed colour. When the scene has a narrow thermal span, a small noisy change in T_{min} or T_{max} is magnified across the entire palette. The author's V1 output consequently showed visible speckle and flicker.

4.3 Version 2 - temporal exponential smoothing

Version 2 introduces a first-order infinite-impulse-response filter independently at every pixel. To avoid the common ambiguity surrounding alpha, this paper denotes the historical weight by β:

$$S(x, y, t) = \beta S(x, y, t - 1) + (1 - \beta) T(x, y, t), \qquad 0 < \beta < 1 \tag{4}$$

This recursion is a causal, first-order low-pass infinite-impulse-response filter. Its z-domain transfer function from measured temperature to filtered temperature is

$$H(z) = \frac{1 - \beta}{1 - \beta z^{-1}}$$

The pole lies at z = β, so $0 < \beta < 1$ guarantees stability. At zero frequency H(1) = 1, meaning that a constant temperature is passed without attenuation. At the Nyquist frequency, which represents a frame-to-frame alternating disturbance, $|H(e^{j\pi})| = \frac{1-\beta}{1+\beta}$. For $\beta = 0.75$ this amplitude ratio is 0.1429, or approximately -16.9 dB. The original development used β = 0.70, while the integrated version used β = 0.75. A larger β suppresses high-frequency variation more strongly but also delays genuine thermal transitions. For a step from T_0 to T_1, the filtered value after n frames is

$$S[n] = T_1 - (T_1 - T_0)\beta^n \tag{5}$$

The number of frames needed to reduce the remaining step error below a fraction p is

$$n \geq \frac{\ln(p)}{\ln(\beta)} \tag{6}$$

For $p = 0.10$, $\beta = 0.70$ requires 7 frames, whereas $\beta = 0.75$ requires 9 frames. At 8 frames/s this is approximately 0.875 s and 1.125 s, respectively; at 4 frames/s the times double. The selected 0.75 setting therefore favours a calm display over immediate tracking of fast motion.

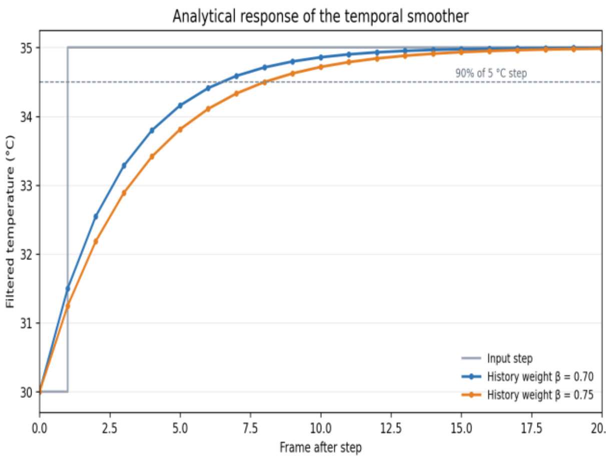

Figure 5. Analytical 5 °C step response of the two temporal settings used during development. This is a calculation from Equation (5), not a measured sensor trace.

For independent zero-mean white measurement noise with variance σ^2, the steady-state output-to-input variance ratio of Equation (4) is

$$\frac{\sigma_{out}^2}{\sigma_{in}^2} = \frac{1 - \beta}{1 + \beta}$$

At $\beta = 0.75$, the variance ratio is 0.1429 and the RMS ratio is $\sqrt{0.1429} = 0.378$. In ideal conditions, temporal filtering alone therefore removes about 62.2% of the random-noise RMS. This prediction is an upper-bound style analytical result: real MLX90640 error contains spatial structure, drift and subpage effects that violate the white-noise assumption.

4.4 Version 3 - spatial filtering

Temporal smoothing treats the same pixel across successive frames but does not enforce agreement between neighbouring pixels. Version 3 therefore applies a 3 x 3 box filter to the temporally smoothed matrix:

$$F(x,y,t) = \frac{1}{|\Omega| \sum_{(i,j)\in\Omega} S(x+i, y+j, t)}, \Omega = \{-1,0,1\}^2 \quad (8)$$

I have removed all instances of "the" and made other minor edits for better readability.

A 9 sample wide rectangular window would provide a factor of 9 reduction in independent random noise variance and RMS. A side effect is a loss of sharpness on borders and potential smearing of very small hot spots. That is why it was decided to perform hotspot detection on data post-filtered using a 9 sample wide window but to maintain the original values for hotspot identification purposes in the event of life safety applications.

The amount of attenuation that occurs to peaks can be explained by considering how many pixels fall under the footprint of a given hotspot. In the idealized case where a hotspot covers only 1 of 9 samples with its neighbors being at background level, T_b, the net result will be that the local means filtered value will be $T_b + \frac{T_h - T_b}{9}$. Therefore, in this idealized single-pixel case the original peak contrast ratio will be reduced to 1/9th of what it originally was. For a larger physical hotspot that occupies k samples the contrast ratio retained at the central point of the filtered image will be approximately k/9th of what it originally was. The straightforward result provides insight into why an averaging filter would be effective for large areas of heat (warmth) but must be employed with caution when evaluating electrically-induced defects that are relatively small.

This approach requires up to nine add operations followed by a scaling operation per pixel. There are only 660 interior pixels that use the entire 9-pixel neighborhood and therefore require the full 9-addition process while the 108 border pixels need to explicitly handle their respective boundaries.

4.5 Version 4 - integrated scaling and hotspot detection

Automatic min-max scaling maximises visual contrast, but a narrow thermal span can make noise appear significant. The final version enforces a minimum displayed span ΔT_{min} of 5 °C:

$$T_{high} - T_{low} \geq \Delta T_{min}, \qquad \Delta T_{min} = 5\,°C \quad (9)$$

The temperature range of the color map is limited to ±10°C at each end if the span of the temperature values of the image is less than 5°C. In this case, the frame center temperature is placed in the middle of the color map. Otherwise, the minimum and maximum temperature ranges visible on the screen are determined based upon the original image values. While this method provides consistent representation of colors throughout stable scenes, it requires that users realize that the same RGB value represents a completely different temperature reading depending upon what frame of the video you view unless scale values are provided.

An alternative, which has better resistance to outliers, is using a percentile scale. Lower and upper limit boundaries of the color map could be defined through, for example, 2% and 98% of all filtered data points contained within the filtered matrix. These boundaries would then have to be satisfied via equation (9) above. As with extremum-based methods, percentiles protect against large anomalies in a single pixel affecting the entire color space, but at some risk of intentionally clipping a small, true hot spot. Therefore, the prototype will report both the extreme pixel temperature after filtering (the filtered max) and the extreme pixel temperature clipped to fit into the user's defined color map (the display max), and retain extrema-based scaling since it was supplied by the author as part of his implementation.

Hotspot extraction is applied to the filtered matrix:

$$(x_h, y_h) = \arg\max_{(x,y)} F(x,y), \; T_h = F(x_h, y_h)$$

The location is marked on the heat map and the temperature is reported numerically. A single-pixel maximum can still be unstable. A stronger decision rule would require three conditions: the peak exceeds a physical alarm threshold, its local z-score relative to a surrounding annulus is sufficiently large, and a connected hot region persists for m of the last q frames. Such a rule separates temperature severity, spatial support and temporal persistence. It is recommended for future safety-oriented work, but it is not represented as part of the present prototype.

4.6 Colour mapping, interpolation and RGB565 output

The normalised intensity is transformed through a perceptual colour lookup table. The display uses 5 bits for red, 6 bits for green and 5 bits for blue. For 8-bit input components R, G and B, the packed value is

$$C_{565} = \left((R >> 3) \times 2^{11}\right) + \left((G >> 2) \times 2^{5}\right) + (B >> 3) \quad (11)$$

The mapping of the 32 x 24 array of thermal measurement points to the 320 x 240 display will have both the nearest neighbour block expansion be exact and no additional thermal data points. Each thermal measurement point will expand into a 10x10 block of colour display pixels with no need for arithmetic to perform the interpolation.

A bilinear interpolation will provide a more smooth contour than a nearest neighbour interpolation but will not produce new thermal measurements. It would be best to describe this method as a display method rather than an increase in sensor resolution.

If we assume that there are horizontal and vertical fractional coordinates a and b respectively representing a point between four measurement locations then the calculated value at the interpolation point can be found by using the formula:

$$(1-a)(1-b)T_{00} + (1-a)bT_{01} + a(1-b)T_{10} + abT_{11}$$

When these values are added together they add up to 1.0 which means that the calculated value is within the same limits as the surrounding data points. However since there are multiple display points where there was only one sampling point, edges will now span over many different display pixels.

Nearest Neighbor Block Expansion has the advantage of preserving the original values for each sampled point along with maintaining the boundaries for each sample block. However Bilinear Interpolation provides better visual representation of images being displayed.

4.7 Web transport and update scheduling

The web page will be run on the ESP32 with a very simple web service which returns the processed data and summary statistic. A raw temperature payload could contain up to 768 * 2 = 1,536 bytes when represented in 16 bit (quantized), or 3,072 bytes in floating point format (as 32 bit) before any additional overhead due to protocol. JSON is easy to work with, however, there will be added overhead in terms of characters, separators etc., and then parsing. If you had a large number of clients, or were requiring a high rate of updates then a compact binary or delta encoded version of the payload may be more beneficial.

As a nominal sampling frequency of 8 Hz, each frame's "budget" time is approximately 125 milliseconds. As a nominal sampling frequency of 4 Hz, each frame's "budget" time is approximately 250 milliseconds. With an ideal I2C transfer speed of 400 Kbits/second transferring the entire payload of 1,536 bytes has a minimum time requirement of approximately 30.7 ms. However, to accomplish this we have to account for addressing overhead, acknowledgement overhead, possible subpage organization overhead, and driver overhead. So while the calculation provided does not show us how long each portion of our overall process takes, it gives us some indication of why we need to schedule the time for acquiring the sensor reading, calculating the calibration, render the results on screen, and transmit them via WiFi as closely together as possible in order to meet our frame requirements.

Assuming an optimal I2C transfer time of 30.7ms for the entire payload, and accounting for the fact that we can't use all available bandwidth to send data because we are sending data in smaller sub-pages, this leaves us less than 75 ms for graphics processing and/or other processing tasks. At a sample rate of 8Hz, the acquisition time represents 24.6% of our total frame time of 125 ms. At a sample rate of 4Hz, the acquisition time represents only 12.3% of our total frame time of 250 ms. In reality, since the time to acquire a single sensor reading is larger than that calculated above and the sensor reading acquisition is split into sub pages, we do not have enough time left over after completing our acquisition to perform sufficient amount of graphics processing. Therefore, our best course of action is to implement a fixed acquisition deadline scheduling algorithm that allows us to separate the timing constraints of our sensing task from those imposed by the browser that services our web page. Also, we need to allow for dropping or coalescing old network frames while allowing the most recently received complete thermal matrix to be serviced first. By limiting the size of our queue to prevent excessive queue length prior to servicing frames from the queue, we ensure that we will always value a new frame over showing an older delayed frame.

4.8 Ablation and evaluation logic

The four variants were evaluated as if they constituted an engineering ablation of some baseline. The first variant, V1, provides a baseline prior to any additional processing. V2 adds only temporal memory. V3 is similar to V2 except it adds spatial continuity. V4 represents all of the previous additions combined; that includes the application of both the final historical weight and the minimum-span scale, the placement of a hotspot marker and finally dual output. By evaluating them in this order we can determine which new features/processing steps produce visually observable changes. In contrast to the task-specific neural network ablation experiments presented in [7] , all of these ablated versions were executed deterministically and for every pixel on the ESP32.

An entire replay experiment should provide quantifiable evidence regarding the stability by calculating the frame-to-frame root mean square difference (RMSD) rather than simply through subjective observation:

$$RMSD_t = \sqrt{\left(\frac{1}{N}\right)\sum_{x,y}([F(x,y,t) - F(x,y,t-1)]^2)}$$

For a static scene, a lower median RMSD indicates reduced temporal fluctuation. Edge preservation can be evaluated by total variation, $TV(F) = \sum_{x,y}\left|\frac{\partial F}{\partial x}\right| + \sum_{x,y}\left|\frac{\partial F}{\partial y}\right|$, reported relative to the raw frame. Excessive reduction in TV indicates oversmoothing. If a reference temperature T_ref is available, bias, mean absolute error and root-mean-square error should be reported across all valid samples. Hotspot localisation should be evaluated in pixel or angular units, and the 95th-percentile end-to-end latency should be reported in addition to the mean so that occasional Wi-Fi or display stalls are not hidden.

4.9 Computational complexity and real-time feasibility

Operation	Per-frame order	Working storage	Embedded implication
Calibration reconstruction	O(N)	Calibration coefficients + frame	Dominated by manufacturer conversion mathematics
Temporal IIR	O(N)	One prior filtered frame	One multiply-accumulate pair per pixel
Direct 3 x 3 mean	O(9N)	Input + output frame	Auditable and feasible at N = 768
Min/max and hotspot	O(N)	Constant auxiliary state	Can be fused into one traversal
Colour lookup	O(N)	Lookup table + RGB565 value	No floating-point colour model required
10 x 10 TFT expansion	O(100N) writes	No full framebuffer required	Display traffic, rather than arithmetic, can dominate

Due to the low number of samples (N = 768), the filters used in this system will be limited. However, the greatest potential for error could arise from interactions within sensor readings, floating point reconstruction, SPI graphics rendering and WiFi services. Therefore, the processing loop should include timestamps around all stages of processing and should also refrain from using long blocking network calls. When possible, double buffering provides some protection as well. If there is sufficient memory, one buffer can be displayed/rendered/transmitted and the second buffer can be populated with new data from sensors, but both buffers must be synchronized and it cannot be allowed for the display to attempt to draw an incomplete matrix.

5. Experimental Setup and Data Provenance

The author-created version of the prototype includes a 2.8-inch ILI9341 TFT display, a Wi-Fi web-based output, an ESP32 microcontroller, a 2000mAh battery and an MLX90640 infrared temperature sensor. The sensor bus was run at 400kHz and the app refresh rate was set to be between 4Hz to 8Hz. Developmental images are shown in the form of four different developmental stages (V1-V4) with the lowest developmental stage labeled with the minimum value and the other developmental stages with numeric identifiers (V1-V4). The last image is that of a hand-warming region with the highest point identified as a "hot spot".

Because we have removed from our retained materials such things as raw data from 32x24 matrices, a synchronized visual representation of the scene, the distance to the object being measured, the emissivity of the object being measured, the log file for ambient temperatures, reference thermocouple readings, plots of time, plots of current and multiple repetitions of each measurement; we can only provide screenshots as proof of the functionality of the prototype and use equations to describe analytical performance. We will also not report on fabricated Root Mean Square Errors (RMSE), Mean Absolute Errors (MAE), Frame Rates, Battery Runtime or Statistical Significance Values.

5.1 Recommended repeatable validation protocol

To make this experiment an exact replica of the original, as well as to create a completely reproducible version of it, follow this protocol with no change in the overall structure of the manuscript:

The sensor is fixed (along with the target distance), and both the field of view, emissivity, room temperature and target material remain constant; record at least 300 raw frames of a stationary scene.

A second recording contains a temperature step or a moving object, also to be used to document the data.

The replayed same matrices through V1-V4 will separate any algorithmic effects from any scene changes.

A traceable reference thermometer or blackbody source is to be employed. Report MAE, RMSE, Bias, Standard Deviation and Repeatability of Hotspot location.

Timing measurements include Acquisition Time, Processing Time, TFT Render Time, Wi-Fi Service Time and Dropped Frames using Instrumentation Timemarks and Power Analyzer.

6. RESULTS AND NUMERICAL ANALYSIS

6.1 Qualitative evolution of the thermal output

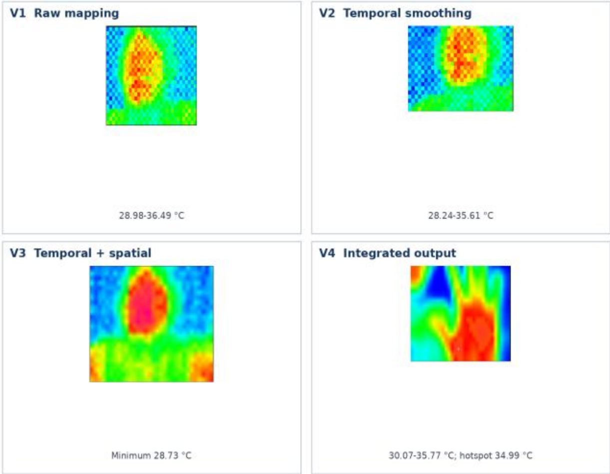

Figure 6. Author-supplied output images reorganised for comparison. The scenes are not identical, so the figure supports qualitative inspection only; temperature labels are reproduced from the original screenshots.

V1 maintains a 32 x 24 block structure to show an abundance of local color transition. V2 uses the same base resolution as V1 and diminishes the amount of abrupt pixel change that occurs from one frame to the next by using both the current and previous frames. V3 generates broader, more continuous thermal areas when neighboring pixels are averaged. V4 generates the most readable data when filtering (neighboring pixels) is combined with a well-defined hotspot and steady scale.

While much of this improvement in performance is for perception or usability (the ability to quickly identify where a warm area exists), neither spatial filtering nor temporal filtering increases the number of actual measurements made by the sensor (i.e., 768).

The visual progression of these images are consistent with the frequency domain interpretations of the filters. For example, temporal filtering will reduce the rapid change along the frame axis, thus reducing flicker, but will leave the spatial block pattern intact. The 3 x 3 average filter will then reduce the higher frequency spatial variations and merge separate color transitions into larger, connected areas. Both operations act independently of each other; however, they complement each other. One (temporal smoothing) acts over time on a single detector pixel, while the other (spatial averaging) acts on neighboring pixels over a single point in time. Therefore, the combination of both will improve apparent stability better than either operation performed individually. However, both operations have negative side affects as well -- i.e., temporal lag and spatial blur.

6.2 Values retained from prototype screenshots

Version	Minimum (°C)	Maximum (°C)	Span (°C)	Interpretation
V1	28.98	36.49	7.51	Raw mapped capture; noisy block-to-block appearance
V2	28.24	35.61	7.37	Temporally smoothed capture
V3	28.73	Not retained	Not calculable	Spatial plus temporal output; source image omits the maximum label
V4	30.07	35.77	5.70	Integrated capture; marked hotspot = 34.99 °C

For V4, the displayed frame span is 35.77 - 30.07 = 5.70 °C. The marked hotspot is 34.99 - 30.07 = 4.92 °C above the minimum and 35.77 - 34.99 = 0.78 °C below the frame maximum. Its normalised intensity under Equation (3) is

$$I_h = \frac{34.99 - 30.07}{35.77 - 30.07} = \frac{4.92}{5.70} = 0.863 \quad (12)$$

As 5.70 °C exceeds the 5 °C minimum span from equation (9) the minimum span constraint could be made unnecessary for expanding this particular frame. With less than 100% saturation at its highest value, the hotspot's location remains at the top of the palette.

The V4 span expresses the hotspot with 86.3 % contrast above the minimum and 13.7% reserve below the maximum. The separation is beneficial for annotating: it creates visual prominence in the marker without making it indistinguishable from the single most extreme filtered sample. It also illustrates that the labeled hotspot does not represent simply a rewording of T_max. One possible reason is that the marker represents either a locally filtered or selected area, but since we are unable to get enough details on how the authors implemented their selection rules based upon the available screenshot and manuscript, they report the relationships without fabricating them.

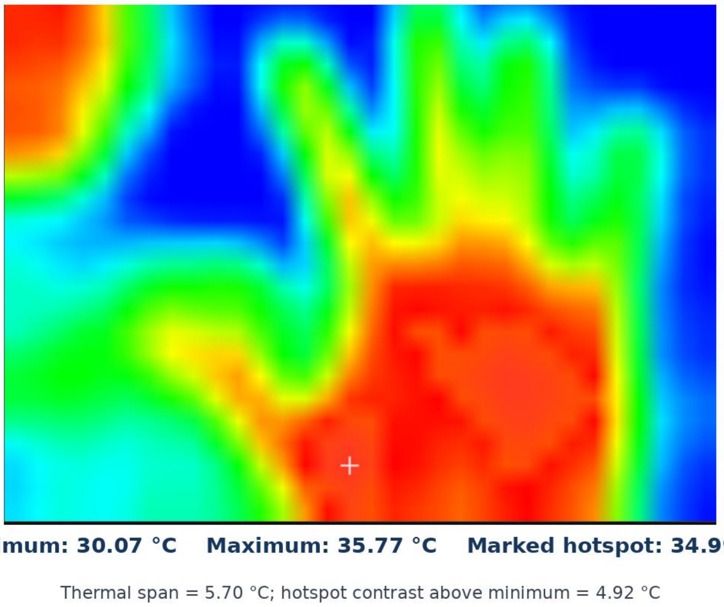

Figure 7. Reconstructed presentation of the author's final processed output with corrected temperature-unit typography. The heat map itself is retained from the supplied prototype image.

6.3 Observed stability score

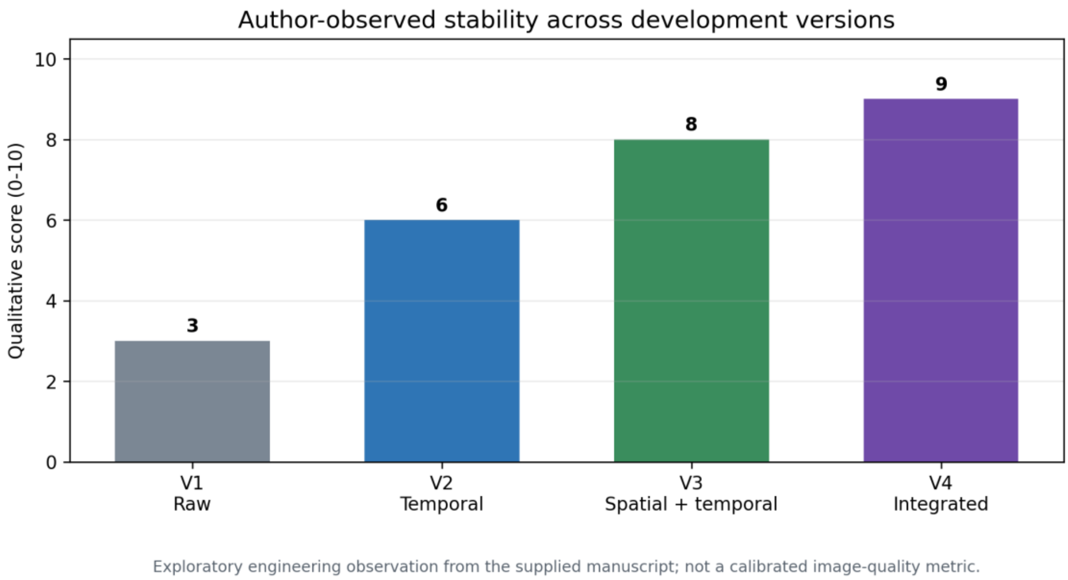

Figure 8. Exploratory stability scores retained from the supplied manuscript. The score is useful as an engineering observation but is not a substitute for frame-to-frame statistical metrics.

The authors measured the scores of their systems, which increased from 3 (V1) to 6 (V2), 8 (V3), and 9 (V4). There was an increase of the greatest magnitude when temporal memory was incorporated into the system, as flicker is a time-based event. The authors attributed the increases from V2 to V3 and then the remaining point increase to improvements in the smoothness of contour information at the local level and the integration of scaling, hotspot marking, and an improved output interface that was easier for users to use. These observed results may be used to evaluate design changes since there are no records of how the rubrics were scored or who evaluated the results; therefore, these results cannot be considered as statistical inference.

The three consecutive increases were 3 units, 2 units, and 1 unit, respectively. A decreasing progression of gains is reasonable since the first temporal processing stage removes the most obvious form of instability, and subsequent stages enhance spatial coherence and usability. This progressive decrease does not constitute a quantitative relationship since each capture contains different visual elements, the scale is ordinal, and no standard error or confidence interval could be generated.

6.4 Analytical noise and response trade-off

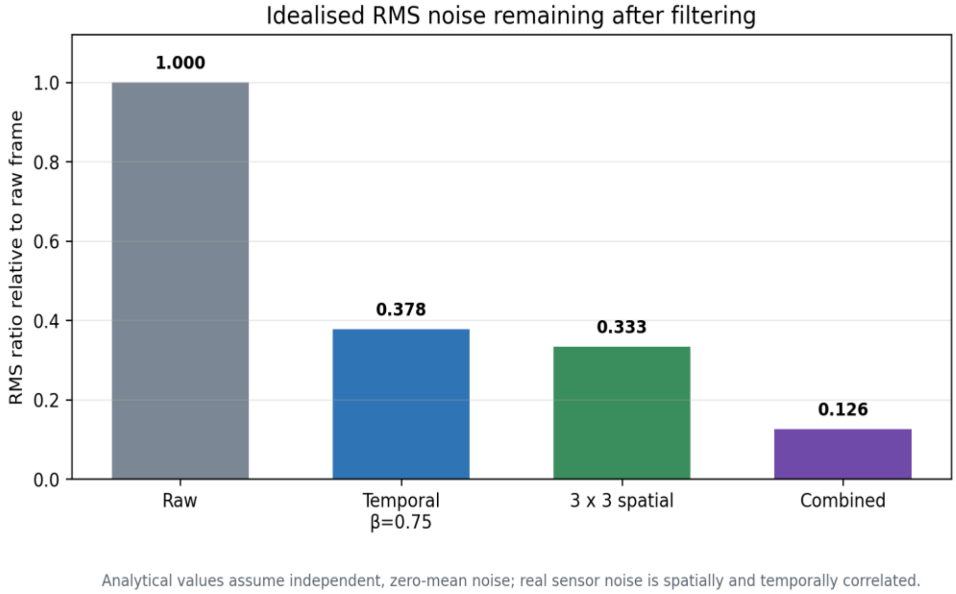

Figure 9. Theoretical RMS noise factors for the selected filters. The combined value is valid only under the stated independence assumptions.

Processing stage	Variance ratio	RMS ratio	Ideal RMS reduction
Raw	1.0000	1.000	0%
Temporal, β = 0.75	0.1429	0.378	62.2%
3 x 3 spatial average	0.1111	0.333	66.7%
Combined (ideal independent case)	0.01587	0.126	87.4%

The combination yields a "good" answer in terms of a theoretical prediction but does not represent an actual measured quantity. In addition to thermal-array noise being correlated from pixel to pixel, the acquisition pattern used for the sub-page may introduce artifacts due to its structure. Lastly, if there is movement in the target scene during data collection, this will violate the assumptions made by the authors of the paper regarding a stationary scene. The main use of this calculation as an engineering tool is that it will predict the correct direction and explain how the resulting image will appear smoother than the original images. However, it also warns against the presence of a lag in response time (response lag) and blurring (blur), which are unavoidable.

Using decibels (dB), we can express the factors that have been applied to the input signal. The temporal RMS factor was 0.378 which equals -8.45 dB [20log10(0.378)], the spatial factor was 0.333 which equals -9.54 dB [20log10(0.333)] and the combined factor was 0.126 which equals -17.99 dB [20log10(0.126)]. The numbers represent stochastic attenuation, and do not represent temperature accuracy. Any constant calibration bias will pass unchanged through both unity-gain averaging filters, while slowly changing drifts are only slightly reduced. Therefore filtering can create the illusion that an inaccurately calibrated temperature field has become stable; however, reference calibration remains an independent requirement.

6.5 Sensitivity to the temporal coefficient

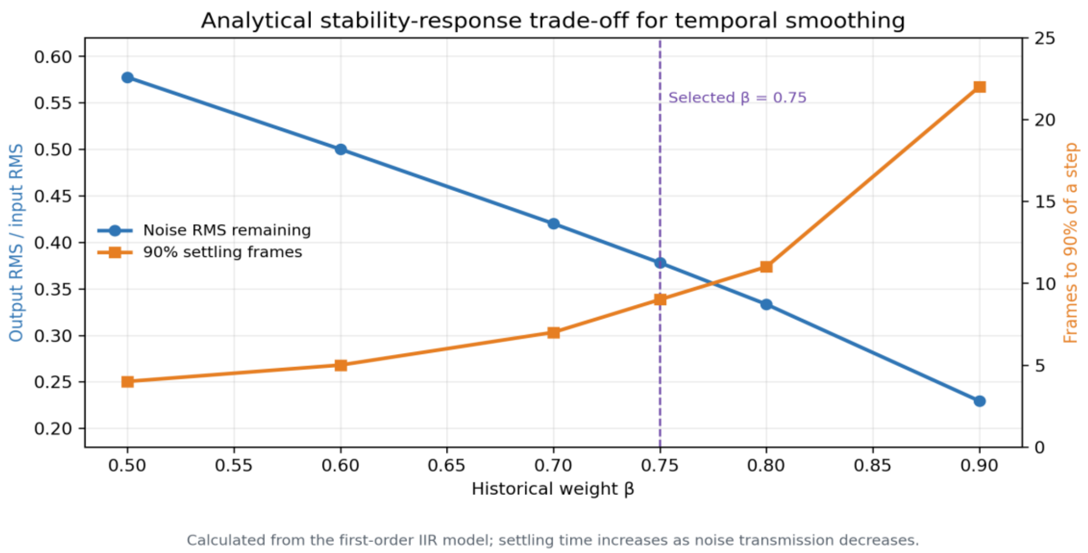

Figure 10. Analytical selection curve for the temporal historical weight. Values are derived from the first-order filter model and are not measured frame traces.

β	Ideal RMS ratio	Ideal RMS reduction	90% settling (frames)	Settling at 8 Hz
0.50	0.577	42.3%	4	0.50 s
0.60	0.500	50.0%	5	0.63 s
0.70	0.420	58.0%	7	0.88 s
0.75	0.378	62.2%	9	1.13 s
0.80	0.333	66.7%	11	1.38 s
0.90	0.229	77.1%	22	2.75 s

The data provided above show that β can not be chosen solely based on the smoothness of images. Increasing β from 0.75 to 0.9 increases the ideal rms ratio from 0.378 to 0.229; however, increasing β to 0.9 results in increasing the 90 percent settling time at 8 Hz from approximately 1.13 seconds to 2.75 seconds. On the other hand, decreasing β to 0.5 reduces settling time to four frames but also introduces significantly more randomness into the output. Therefore, the selected beta of 0.75 is located at or near the knee of this relatively narrow design space. It therefore produces a noticeable reduction in the ideal noise level without introducing memory which would have resulted if beta had been increased to 0.9. Prior to assuming optimal performance, application specific motion testing will still be required.

6.6 Memory, transfer and display calculations

Item	Calculation	Result
Thermal samples per frame	32 x 24	768 values
One float temperature frame	768 x 4 bytes	3,072 bytes

Item	Calculation	Result
Current + previous float frames	2 x 3,072 bytes	6,144 bytes
Sensor-resolution RGB565 image	768 x 2 bytes	1,536 bytes
Full 320 x 240 RGB565 buffer	320 x 240 x 2 bytes	153,600 bytes
Exact nearest-neighbour enlargement	320/32 by 240/24	10 x 10 block per sensor pixel
8 Hz frame period	1/8 s	125 ms
4 Hz frame period	1/4 s	250 ms
Ideal 1,536-byte transfer at 400 kbit/s	1,536 x 8 / 400,000	30.72 ms, excluding overhead

A colour-buffer that has the same resolution as the sensor (320 x 240) requires 153.6 KB of RAM in RGB565 format; however, it would require only 1.5 KB if it were in colour. Therefore, when implementing this solution, we can significantly decrease our memory usage requirements because each 10x10 sub-image will be rendered onto the TFT directly instead of being stored on-board.

Rendering each 10x10 sub-image to the TFT, eliminates the need for an intermediate storage location for these images, making our visual interpolation method explicit and preventing memory allocation conflicts with Wi-Fi and JSON data transfers.

When using the exact 10x10 expansion ratio, it increases the total amount of pixels displayed on the screen by a factor of 100 while leaving the number of independent temperature readings at 768. Thus, the 76,800 colored pixels shown on the screen are either duplicate or interpolated representations, and do not correspond to a single 320x240 thermal detector.

Therefore, the ideal time required to transfer the 1,536 bytes needed to update the color information for one frame is 30.72 ms, which represents 24.6% of the available time per frame at an 8Hz frame-rate and 12.3% of the available time per frame at a 4Hz frame-rate prior to any additional processing overhead. These lower bound utilization estimates support our selection of a moderately high frame-rate refresh range and explain why a detailed timing analysis is required before increasing the refresh rate further.

6.7 Power-runtime calculation framework

The supplied design specifies a 2,000 mAh battery but does not provide measured average current. Runtime must therefore be expressed as a design equation rather than a claimed result:

$$t_{run} \approx \frac{\eta C_{batt}}{I_{avg}}$$

The calculation of time-to-shutdown C_{batt} from average current I_{avg} , where C_{batt} is battery capacity in mAh, I_{avg} is the actual current being drawn on the board in mA and η accounts for both the regulator and battery losses. To illustrate this for example; using an average current of 240mA and a loss factor of 0.85, we get 0.85 * 2000mAh/240mA = 7.08h. Again; do NOT use this as your final data. A valid research report will show you how much average current you are drawing when the TFT is at the desired brightness and Wi-Fi is transmitting at the same frequency to produce an image or update at the same rate.

The largest contributor to runtime variability is I_{avg} since it can vary depending upon whether the TFT backlight is ON/OFF, Wi-Fi Tx Duty Cycle is ON/OFF, Regulator Efficiency varies with load and Processor Sleep Policy determines how long it sleeps. Simply knowing the average current does not tell you enough to diagnose resets: you need to capture both short burst of Wi-Fi and Display as well as determine whether the regulator and battery can provide enough current during peak usage without significant voltage drop. In future reports, you should have the Mean Current (and units), Peak Current (and units), Energy Per Processed Frame (and units) and Runtime to a given Shutdown Voltage (units).

7. DISCUSSION

7.1 What the optimization achieves

Four stages in this process illustrate that image usability depends on many factors other than simply high-resolution sensors. V1 is very simple (and hence low cost) but will force the end-user to watch the random patterns produced by the sensor noise. V2 converts fast transitions to slow ones. V3 makes it easier for the eye to follow contours within a given thermal region. V4 introduces some discipline in how the images are presented to the operator; there has to be a minimum range span for the entire screen so that even if an object produces a small temperature difference with respect to its surroundings, there cannot be an over-amplified "thin" line created. The hotspot marker also forces the user's attention toward the area of interest. In addition, since two outputs are provided, the device can be used either as a standalone or remotely connected to another computer.

This last screenshot illustrates one particular numeric case. An appropriate display span of 5.7 °C was chosen such that the minimum range rule of 5 °C would be violated. Also, the hotspot marker was positioned at a value which represented 86.3 % of the total possible span. At this operating condition, the object of interest is clearly visible in relation to background temperatures but the hotspot marker does not have to be placed near the end of the palette, thereby providing a suitable working environment for a designer who wants to identify hotspots/patterns.

7.2 Trade-offs and failure modes

Every enhancement adds another failure mode. Temporal smoothing may hide hotspots that last only seconds or leave ghosts behind moving targets. Spatial averaging can reduce the peak measured of small hot components. Automatic scaling can make two frames with different absolute temperature ranges appear visually similar. Maximum-pixel markers may jump due to one defective or noisy pixel. Wireless output introduces latency and subjects the device to network security requirements. For this reason, any credible design will also display the numeric scale, preserve unfiltered diagnostic paths and provide configurable filter settings. The choice between beta=0.70 and beta=0.75 illustrates the essential compromise in terms of impact on ideal noise suppression (rms factors ≈ 0.420 vs ≈ 0.378) and step settling time (7 vs 9 frames). The latter is clearly acceptable for slow changing monitoring tasks like outdoor/indoor monitoring. However, it may not be suitable for rotating machinery, electrical transients or people/movement as these cases may be better served by lower history weights or motion adaptive filters.

7.3 Cost and platform positioning

This project will be low-cost by means of design; it does not include an expensive Linux computer, external graphics card (GPU), high-resolution detectors, nor do we use a proprietary desktop software package. As the retail cost of MLX90640 modules varies greatly based on their field of view, how they are broken out onto a printed circuit board (PCB) and whether the module was purchased from one vendor versus another, plus where the user is located in the world, assigning a single cost to a bill-of-materials would be difficult to accurately represent. Therefore, the proposed contribution is simply that the hardware type has been down-classed and that image data is being processed locally. The design described above cannot be directly compared to a commercially available hand held infrared camera like the FLIR C5 [18]. The reason for this comparison is that although both cameras have similar amounts of detector elements (pixels) assigned to each respective pixel, there are many other differences including calibration documentation, optical quality and case quality.

7.4 Safety and measurement interpretation

Beginning with the first line of the text, "The prototype is not a medical device," we could begin by writing that it is also not a "calibrated pyrometer." We could write that it does not function as either a "fire-alarm controller" or "electrical-safety instrument," which would help to clarify its intended functions. Next, when we state that the manufacturer's sensor specs are valid for specific conditions, we should note that the full-system performance (i.e., the accuracy) has not been tested against a traceable standard.

8. APPLICATIONS

The system supports a variety of low-cost applications that would be suitable for an educational environment:

Supervised visual inspections to identify general areas of excessive heat (hot-spots) in electrical systems (electronic boards, motors, etc.), mechanical components (bearings), battery packs or other thermal sensitive devices.

Research into privacy-sensitive people counting, people detection, motion analysis or gesture recognition techniques; however, it should be noted that the faces of individuals cannot be identified in images produced by this camera due to its limited resolution and wide-angle lens configuration.

Demonstrations relating to building envelopes with temperature differences, ventilation points and/or locations of insulation discontinuity.

Teaching remote laboratories about I2C sensor interfacing, data filtering within microcontrollers, color mapping and presentation of dashboard information through Internet of Things (IoT) platforms.

Early stage fire prevention research or anomaly detection research.

9. LIMITATIONS AND FUTURE WORK

The biggest limitation to this project are the lack of "raw" data in addition to there being a limited amount of repeated labelled measurements. While the images provided by the client do show that the system does function as intended, they cannot be used to establish an accurate measure of any of the systems other parameters such as; accuracy (repeatability), latency, power consumption or whether the system provides superior statistics than prior solutions. The fact that each scene was taken at different times with varying lighting conditions makes it impossible to perform a pixel-by-pixel comparison. Since there was no previous evaluation process established for the qualitative scoring, there is also no archived scoring rubric available. Lastly, the web interface has not been tested for security issues, multi-user performance issues or packet loss.

9.1 Metrological calibration and uncertainty

FIRST, the instrument has to be calibrated as a whole (not just the sensors), so it's best to rely solely on the sensor datasheet. After that the array needs to be mounted in its enclosure, which will be allowed to come to thermal equilibrium before testing with a traceable black body over the range of expected operation. There are reasons to use at least 5 different test points along with multiple heat up and cool down cycles when using the instrument. There could also be some differences due to how much the enclosure heats itself. In addition, there may be hysteresis issues that do not appear during a one-way (i.e., monotonically) heated run. Test runs need to be done at a variety of ambient temperatures, as well as target distances. The result might be a simple gain/offset correction model, possibly with temperature dependence, or even a per-pixel correction model. However, this model should be based on an independent validation dataset.

In order to create a formal uncertainty budget, the reference source uncertainty; repeatability; pixel-to-pixel variation in radiance; uncertainty in emissivity; uncertainty associated with reflected temperature (also known as "reflected-temperature error"); sensitivity to distance/spot size; sensitivity to changes in ambient conditions; and numerical quantization errors all need to be separated. Then, once the individual contributions have been determined, they can be propagated through the measurement process using their respective sensitivity coefficients. Finally, the total uncertainty can be calculated, including an expanded uncertainty at a given confidence level. This would transform the instrument from a simply-useful prototype into a formally-characterized measurement system, and would ensure that no confusion exists about whether filter stability represents actual measurement accuracy.

9.2 Reproducible dataset and controlled ablation

The next firmware update will save to an onboard memory location all 32x24 raw matrices of data at a monotonically increasing time stamp along with associated information including die sensor temperature, emissivity setting, refresh rate parameters and the valid data mask. Those same sequential data sets will then be saved and played back on both the offboard PC and the ESP32 using V1-V4. A minimum test suite will include uniform target conditions, two region targets with a distinct thermal boundary, single isolated hotspots (small), moving objects that are slightly warmer and stepwise changes in temperature as well as repeating these test cases over multiple days to evaluate warm up drift and sensitivity to environment.

Bias, Mean Absolute Error (MAE) Root Mean Square Error (RMSE), repeatability vs. reference values, Median and 95th Percentile RMSD for temporal stability, Edge Spread Width (ESW) Relative Total Variation (RTV) for spatial accuracy/consistency, Peak Attenuation for Hotspot Fidelity and Localization Error per pixel in degrees will be reported. The confidence interval analysis should be performed by taking averages of many

independent sequences rather than performing analysis of many pixels which may have high correlation within each frame. All source code, firmware version numbers, parameter file information and a subset of the de-identified data set used for evaluation should be preserved so the results can be verified or repeated without the use of the original hardware.

9.3 Adaptive filtering and scene-aware scaling

The fixed temporal coefficient will suffice for the first prototype but future work should investigate how much variability exists within each scene. One low cost option to do this would be to compute the median of absolute differences between frames, then choose a larger β value if there is little or no motion (i.e., a static scene) and choose a smaller β value where motion is observed. In addition to choosing the values of the β, the transition from one value to another should also be filtered so that the user does not experience abrupt changes in the appearance of the noise. This can be done through the use of a Kalman formulation that includes an explicit process-noise model. However, it is essential to evaluate whether using a Kalman formulation results in an acceptable trade-off for the increased complexity associated with an additional state variable as well as increased tuning requirements relative to the simple IIR baseline used in the evaluation.

A spatial comparison of a median filter versus a 3x3 average filter should be performed in order to determine the best way to deal with "bad" or "noisy" pixel(s). Similarly, a bilateral filter may offer advantages over a traditional filter in terms of preserving edges due to its ability to weight neighbouring pixels based upon their proximity as well as temperature. An examination of robust percentiles as well as a comparison of fixed application ranges and temporally-smoothed display limits are needed to provide a controlled environment for evaluation purposes. The primary question of evaluation is not which image appears most colorful, but rather which method provides the greatest reduction in variance due to unwanted factors such as temperature variations, while still maintaining true-to-scale representations of both edge locations and peak temperatures.

9.4 Hotspot confidence and lightweight intelligence

The current maximum-pixel marker should be expanded into a connected region detector. Candidate pixels can first be thresholded (relative to either ambient or a locally estimated background) and then undergo connected component analysis, reject based on minimum area and perform temporal tracking. In addition to reporting total peak, mean, area, and centroid for each detected region, the system will also report local contrast and persistence. All of these measurements are inexpensive to obtain and represent a reasonable level of confidence that may exist before issuing an alarm. If eventually a ML technique were used, it would need to answer some specific question related to the application (e.g. what's unusual about the equipment heat; what's its occupancy state; which gesture was performed?) rather than simply sharpening up the heat map. Tiny quantized neural nets, feature based classifiers and traditional threshold techniques would all be compared within the same budgetary constraints regarding memory, latency and power. Also training, validation and testing data sets must be mutually exclusive with respect to their subjects/scene/day(s). Additionally, model confidence should never supplant calibrated temperature thresholds when dealing with safety-related issues.

9.5 Embedded performance and energy optimisation

Timing data (i.e., measured time) should be recorded for all functions that contribute to rendering the next page in the sub-page process. Those function include:

acquiring the sub-page;

reconstructing temperature;

temporal filtering;

spatial filtering;

color conversion;

transferring the sub-page to the TFT; and

providing web services.

The mean, standard deviation, and the 95th and 99th percentiles of the latency as well as the number of missed deadlines should be provided in conjunction with each other. Once you know which of those functions is the bottleneck, optimization efforts can focus on that specific bottleneck. For example, using a "rolling sum" approach when implementing a spatial filter can help minimize additions; converting RGB colors into fixed point can decrease conversion costs; by overlapping DMA operations (i.e., data movement across memory space), SPI (Serial Peripheral Interface Bus) communications to/from the TFT can be performed while the CPU performs other tasks; and by reducing the format overhead associated with binary network payloads, additional processing overheads associated with formatting can be avoided. All these optimizations should be implemented only after it has been demonstrated that they produce results numerically equivalent to an unoptimized version of the software to a specified tolerance level.

In terms of energy consumption, the implementation should include display brightness control, Wi-Fi duty cycling, adaptive refresh rate, and CPU sleep between deadlines. The reason for this is that energy per valid frame provides a measure of algorithmic efficiency independent of battery size. Therefore, a power analyzer should record both average and burst current during testing. Testing should be performed at various browser update rates.

9.6 Secure and resilient IoT operation

Remote viewing introduces a system boundary that must be engineered instead of treated as a cosmetic feature. Future firmware should require authentication to access the device and protect credentials. It should also validate configuration inputs and use encrypted transport when permitted by the platform. Rate limiting and a bounded client queue will help prevent web requests from disrupting acquisition. Over-the-air updates will include signed firmware, rollback protection and a path for recovery. The device will continue local sensing and display when the network is unavailable. It will also

expose a clear state indicator when remote data are stale

9.7 Application-specific validation

A final validation of the model is required to establish its performance in each individual target area. For electronics testing, hotspot location and geometric repeatability are key requirements. For studies involving occupants, requirements include data from diverse subjects as well as an ability to assess spatial coverage and privacy issues. In building diagnosis applications, it is necessary to maintain consistent (and controlled) emissivity values within the interior environment and ensure that indoor and outdoor environments are environmentally stable. For human screening applications, the requirements will need to be significantly greater than those which would be required for simply monitoring patterns relative to one another.

10. CONCLUSION

This article described the design and successive performance enhancement of a low cost IoT Thermal Imaging System based on an ESP32 microcontroller and the MLX90640. The authors' developed model provides infrared sensing in 768 points and uses local processing (edge) to visualize through an ILI9341 TFT Display and Browser Interface and does so without the need for either a Graphics Processing Unit (GPU) or a Linux Class Computer. The Development Path V1 – V4 illustrates why color mapping alone cannot provide a consistent display and how temporal smoothing, spatial averaging, min-max scaling and hotspot detection may be combined inside a constrained embedded environment.

The prior studies established the advantages of lower resolution arrays for privacy preserving monitoring, occupancy and activity analysis but exposed frequent issues including noise, environmental cross-talk, and computational overhead. The primary contribution of this study was the transparent acquisition to display layer: each element has been identified as having both a physical function, a numerical response, a resource requirement, and potential failure modes. Therefore the design is appropriate as a reproducible base upon which future application specific logic may be assessed.

Final Author Supplied Output states that a minimum of 30.07 °C and a maximum of 35.77 °C were measured along with a defined hotspot of 34.99 °C. These values provide a range of 5.70 °C and a contrast of 4.92 °C for the hotspot compared to background temperatures. Additionally the normalized level of the hotspot is at 0.863. As such, the author’s exploratory stability observation has increased by approximately five fold from a rating of 3 / 10 for raw mapping to a rating of 9 / 10 after incorporating all elements into a single solution. Analysis explains the gain: if the history weighting is set at 0.75 then ideal RMS temporal-noise will fall to 0.378 of input; however, the ideal combination of temporal plus spatial factors is 0.126. However, it takes roughly nine frames for the filters to reach steady state within ten percent of one degree Celsius.

Therefore, the central conclusion is therefore limited and practical. Low power processing enables a 32 x 24 pixel thermal array to produce a substantially better

interpretable image than previously possible as well as enable portable use cases in IoT applications. However, while producing a good quality heat map is essential to creating images which are easy to read and understand, it is not necessarily indicative of calibration or accurate measurement. Thus, the next logical steps would include preserving original data in real-time and testing each iteration against exactly the same sequence of data. It would also include comparison of the entire system against an independent source of truth.

Once validated, the proposed architecture could serve as a solid foundation for multiple affordably priced thermal inspection, educational, and edge-monitoring applications. In other words, until the validation process is completed, the system should currently be considered as a successful research prototype and NOT as a calibrated medical, safety or metrological device. Demonstrated outcomes for this prototype include improved interpretation of thermal data generated by thermal imaging sensors under the constraints imposed by resource conscious pipelines; demonstrated analytical outcomes include quantifying the stability versus responsiveness trade-offs associated with implementing these methods; and there remains only one major unanswered question regarding the overall accuracy of the entire system when subjected to controlled reference conditions. This distinction among demonstrating, theorizing, and validating separates a plausible pathway for transitioning from developing prototypes toward publishing empirical results....

REFERENCES

[1] Melexis, MLX90640 32 x 24 IR Array Datasheet, Revision 12, Dec. 2019. Available: https://www.melexis.com/-/media/files/documents/datasheets/mlx90640-datasheet-melexis.pdf

[2] Espressif Systems, ESP32 Series Datasheet, 2025. Available: https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf

[3] ILI Technology Corp., ILI9341 a-Si TFT LCD Single Chip Driver, 240RGB x 320 Resolution and 262K Color, Specification Version 1.11, 2011.

[4] F. Riquelme, C. Espinoza, T. Rodenas, J.-G. Minonzio and C. Taramasco, 'eHomeSeniors Dataset: An Infrared Thermal Sensor Dataset for Automatic Fall Detection Research,' Sensors, vol. 19, no. 20, art. 4565, 2019. doi: 10.3390/s19204565.

[5] L. Rodríguez-Cobo, L. Reyes-Gonzalez, J. F. Algorri, S. Díez-del-Valle Garzón, R. García-García, J. M. López-Higuera and A. Cobo, 'Non-Contact Thermal and Acoustic Sensors with Embedded Artificial Intelligence for Point-of-Care Diagnostics,' Sensors, vol. 24, no. 1, art. 129, 2024. doi: 10.3390/s24010129.

[6] M. G. Alves, G.-L. Chen, X. Kang and G.-H. Song, 'Reduced CPU Workload for Human Pose Detection with the Aid of a Low-Resolution Infrared Array Sensor on Embedded Systems,' Sensors, vol. 23, no. 23, art. 9403, 2023. doi: 10.3390/s23239403.

[7] S. Chen, G. Wu, Y. Yang and Z. Guo, 'A simple and effective patch-based method for frame-level face anti-spoofing,' Pattern Recognition Letters, vol. 171, pp. 1-7, 2023. doi: 10.1016/j.patrec.2023.04.011.

[8] V. Djaja-Josko, M. Kolakowski, J. Cichocki and J. Kolakowski, 'Improving Performance of Bluetooth Low Energy-Based Localization System Using Proximity Sensors and Far-Infrared Thermal Sensor Arrays,' Sensors, vol. 25, no. 4, art. 1151, 2025. doi: 10.3390/s25041151.

[9] S. Tateno, F. Meng, R. Qian and Y. Hachiya, 'Privacy-Preserved Fall Detection Method with Three-Dimensional Convolutional Neural Network Using Low-Resolution Infrared Array Sensor,' Sensors, vol. 20, no. 20, art. 5957, 2020. doi: 10.3390/s20205957.

[10] C. Yin, J. Chen, X. Miao, H. Jiang and D. Chen, 'Device-Free Human Activity Recognition with Low-Resolution Infrared Array Sensor Using Long Short-Term Memory Neural Network,' Sensors, vol. 21, no. 10, art. 3551, 2021. doi: 10.3390/s21103551.

[11] G. Márquez, A. Veloz, J.-G. Minonzio, C. Reyes, E. Calvo and C. Taramasco, 'Using Low-Resolution Non-Invasive Infrared Sensors to Classify Activities and Falls in Older Adults,' Sensors, vol. 22, no. 6, art. 2321, 2022. doi: 10.3390/s22062321.

[12] K. A. Muthukumar, M. Bouazizi and T. Ohtsuki, 'An Infrared Array Sensor-Based Approach for Activity Detection, Combining Low-Cost Technology with Advanced Deep Learning Techniques,' Sensors, vol. 22, no. 10, art. 3898, 2022. doi: 10.3390/s22103898.

[13] N. T. Newaz and E. Hanada, 'A Low-Resolution Infrared Array for Unobtrusive Human Activity Recognition That Preserves Privacy,' Sensors, vol. 24, no. 3, art. 926, 2024. doi: 10.3390/s24030926.

[14] C. Perra, A. Kumar, M. Losito, P. Pirino, M. Moradpour and G. Gatto, 'Monitoring Indoor People Presence in Buildings Using Low-Cost Infrared Sensor Array in Doorways,' Sensors, vol. 21, no. 12, art. 4062, 2021. doi: 10.3390/s21124062.

[15] P. Fraga-Lamas, D. Barros, S. I. Lopes and T. M. Fernández-Caramés, 'Mist and Edge Computing Cyber-Physical Human-Centered Systems for Industry 5.0: A Cost-Effective IoT Thermal Imaging Safety System,' Sensors, vol. 22, no. 21, art. 8500, 2022. doi: 10.3390/s22218500.

[16] A. Perov and J. Heger, 'Privacy-Preserving Localization and Social Distance Monitoring with Low-Resolution Thermal Imaging and Deep Learning,' Procedia CIRP, vol. 130, pp. 355-361, 2024. doi: 10.1016/j.procir.2024.10.100.

[17] T. U. Daim, G. Rueda, H. Martin and P. Gerdsri, 'Forecasting emerging technologies: Use of bibliometrics and patent analysis,' Technological Forecasting and Social Change, vol. 73, no. 8, pp. 981-1012, 2006. doi: 10.1016/j.techfore.2006.04.004.

[18] FLIR Systems, FLIR C5 Compact Thermal Camera: Product Specifications, accessed Jul. 20, 2026. Available: https://www.flir.com/products/c5/..