

Infrared dataset generation for people detection through superimposition of different camera sensors

Alessandro Avi, Matteo Zuccatti, Matteo Nardello, Nicola Conci, and Davide Brunelli

{*alessandro.avi, matteo.zuccatti*}@*studenti.unitn.it*, {*nicola.conci,davide.brunelli*}@*unitn.it*

University of Trento, Via Sommarive, 9, 38123 Povo, Trento TN

Abstract

Infra-red (IR) cameras have found widespread use in many different fields. The most common ones are generally related to industrial applications, particularly maintenance and inspections activities. In the domain of surveillance, instead, they are mostly used for threat detection and security purposes. Pushed by cost reduction and the availability of compact sensors, intelligent IR cameras are gaining popularity in the field of Internet-of-Things, in light of the valuable information made available by the acquired data. Unfortunately, the achievable overall quality is not always satisfactory. For example, low-resolution devices, noise, or harsh environmental conditions, like high temperatures on sunny days, can degrade the quality of the thermal images. This paper presents the development of a portable, low-cost, and low-power thermal scanner prototype consisting of a thermal sensor assisted by a grayscale camera. The prototype is completely made using COTS components and provides 80×60 IR and 160×120 grayscale images, mostly used to collect and validate the IR-based data. Our application focuses on people detection, for which we present a suitable learning framework together with the corresponding IR dataset, collected and annotated via the paired grayscale images.

1. Introduction

In recent years with the increased interest in security and surveillance applications, advanced driver assistance systems, autonomous vehicles, and human behaviour analysis, people detection has become a key research problem in the computer vision research community [1, 2, 3]. Despite the many successful approaches presented in literature, which guarantee very promising performance on the existing benchmark datasets, the in-the-wild still detection might still present challenges. Appearance variations, illu-

mination variations, occlusions, human motion variations, and background noises are some of the challenges currently faced by the research community; such troubles become even more evident when the available computational resources are limited, as in the case of embedded systems. To ease this task, researchers have proposed to fuse thermal imaging and a traditional camera to detect human subjects [4]. Thermal imaging, also referred to as InfraRed Thermography (IRT) is a technique that detects the intensity of radiation in the infrared part of the electromagnetic spectrum and visualizes the recorded temperatures as 2-D images. It has the advantages of providing high-precision and non-invasive temperature measurement as a non-destructive test method to detect temperature-related features. Thermal imaging sensors have been adopted in various environments to detect human subjects and heat signatures [5, 6, 7, 8].

The recent literature shows that efforts still focus on high-resolution thermal imaging. Nevertheless, thanks to technological progress and the availability of low-cost and low-power thermal sensors, low-resolution thermal imaging has emerged as trending research topic [9, 10, 11, 12].

This paper presents a portable and low-cost thermal scanner prototype, capable of acquiring both grayscale and IR images through two separate sensors. To demonstrate the feasibility of the solution, the current work focuses on using the sensors to perform human detection. We provide a labelled dataset containing 3765 pairs of IR and grayscale images. Compared to other datasets, the one we propose is suitable for consumer implementations with uncalibrated and low-resolution devices. Despite the small image size, we show that it is still possible to rely on existing learning frameworks to train a reliable detector.

To achieve the above-mentioned goal, we developed a thermal scanner prototype board together with the relevant software, interfacing the microcontroller (a STM32F4 MCU), with the two camera modules; a Lepton Flir v2.0 thermal sensor and a Himax HM01B0 grayscale VGA sen-

sor. Further details about the chosen hardware are provided in the next sections.

The acquired images are then transferred to a PC, where the training of the people detection module is performed off-line. In order to train the model for people detection, an ad-hoc annotation strategy has been devised, superimposing the visible-spectrum and IR data, and using a common state-of-the-art detector to create the corresponding bounding boxes. Further details about each processing stage are reported in the coming paragraphs. Figure 1 shows the developed prototype.

The main contributions of this paper are:

- The development of a portable thermal scanner based on a STM32F4 MCU with two camera sensors. A Lepton Flir IR camera and Himax HM01B0 grayscale camera;
- The development of an image pre-processing algorithm to correctly align and fuse IR and RGB images
- The collection and annotation of a publicly available dataset;
- The training and fine-tuning of a YOLO [13] detector for people detection in the IR domain;

The remainder of this paper is organized as follows: the relevant related work is reviewed in Section 2. Section 3 describes the hardware used and the software developed. In Section 4 we present the acquired dataset, while the detection algorithm is discussed in Section 5. Section 6 describes the simulations and experimental results, also comparing against the existing state-of-the-art. Section 7 concludes

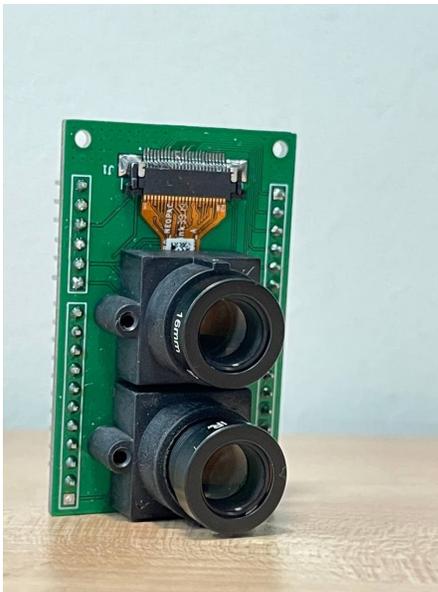


Figure 1: Thermal scanner prototype.

this article with some final remarks and some considerations regarding potential future work directions.

2. Related Work

In recent years, thermal imagery [6, 7, 8] has become a trending topic thanks to its wide applicability in different sectors. Possible applications span from the agriculture sector [14, 15] to the industrial sector [16] and also to the medical [17] domain. The authors in [18] describe how a drone equipped with an IR camera has been used as an effective solution for soil monitoring. Soil temperature, in fact, is closely related to the ability of the terrain to retain water, thus determining its fertility. Different research studies can be found in the literature with regard to the industrial sector. In [19, 20] the authors describe how a UAV equipped with an IR camera can be a viable solution for the inspection of photovoltaic installations. In [21] an IR camera is exploited to enhance pedestrian visibility. The authors point out that, although the information acquired by the IR camera is insightful, a standard grayscale camera should also be added. For example, on bright sunny days, the temperature of the road increases, making the silhouette of the pedestrians less detectable in the IR spectrum. In those cases, the grayscale camera works at its peak performance, with a minimum signal-to-noise ratio (SNR). Therefore, the paper proposes an advanced solution based on deep neural networks to optimally fuse the image from the two visual sensors to achieve better pedestrian recognition accuracy.

Machine learning applied to thermal imagery, is becoming the standard de facto for thermal image analysis and manipulation. In the literature, the topic is well investigated by different research works. In [22] pedestrian detection is implemented by evaluating thermal images through a CNN. In [23] and [24] infrared thermography is used for automatically detecting anomalies using deep neural networks (DNNs). In [25] neural networks are used to classify objects in thermal images for search and rescue missions using UAVs. Deep neural networks are exploited yet again in [26] for fast eye tracking from thermal images. Finally, infrared images are also being used to enhance face recognition applications [27]. The importance of thermal imagery is also confirmed by the continuous increase of publicly available datasets for training learning models, as presented in Table 1, where we report the details of some freely-available datasets.

Fueled by recent technological advantages, neural networks are being also used to enhance the performance of automatic analysis of low-resolution thermal images [36, 37, 38, 39, 40, 12, 10, 11]. To ease this task, more recent works are trying to fuse images from multiple sources into a comprehensive image [41, 42, 43, 44]. As reported in [45], combining multiple images obtained by different kinds of sensors generates a robust and informative image that can

Table 1: Example of thermal imaging datasets freely available online

Dataset Description	Ref
L-CAS Thermal Physiological Monitoring Dataset ~ 3000 382 x 288 pixels images.	[28]
A Face Database Simultaneously Acquired in Visible, Near-Infrared and Thermal Spectrums, ~ 7380 640 x 480 pixels images.	[29]
KAIST Multispectral Pedestrian Dataset, $\sim 95k$ 640 x 480 color-thermal pairs images	[30]
Far-infrared human action dataset, $\sim 75k$ 16 x 16 pixels images	[31]
Large-Scale High-Diversity Thermal Infrared Object Tracking Benchmark, $\sim 600k$ images	[32]
LWIR Thermal Imaging Dataset for Person Detection, $\sim 12K$ 1280 x 960 pixels images	[33]
FLIR Thermal Dataset for Algorithm Training, $\sim 18k$ 640 x 512 pixels images	[34]
CASIA Infrared Night Gait Dataset, 129 x 130 pixels images	[35]

facilitate the training phase of a learning architecture.

3. System Architecture

The system we propose in this paper is designed and conceived to create a compact and portable thermal scanner suitable for collecting a thermal image dataset and validate the potential of low-resolution IR images. The main components of the device are a microcontroller, two camera modules (visual and IR) and a 3D printed camera support with 2 degrees of freedom, allowing the correct alignment of the two visual sensors, which have different fields of view and to guarantee an easy customization of the setup depending on the specific needs. The collected images are then streamed using a serial connection to a nearby computer. The components used for building the thermal scanner are presented in Figure 2.

MCU. The thermal scanner is built around a STM32F4¹ high-performance MCU. To ease the prototyping phase, we chose to use the Nucleo-F401² development board, providing a wide range of connectivity protocols (i.e., I2C, SPI, UART). Finally, the integrated programmer and available code sample allow a straightforward development of the streaming software.

IR sensor. For the IR camera module, we used a Flir Lepton³ camera module mounted on the Breakout Board v2.0. One of the advantages of this device is having both I2C and SPI peripherals. The I2C protocol is used to set up and properly configure the camera settings, while the SPI

takes care of the data transmission. The resolution of the Lepton Flir is 80x60 pixels. The horizontal field of view of the camera is 50° and the scene dynamic range spans from -10°C to 140°C. The device will capture either black and white (grayscale) or coloured images depending on the camera settings. In the latter case, a few color palettes are already built-in, with the option for customizations. Another useful built-in feature is the Automatic Gain Control (AGC), through which the dynamic range of the sensors is remapped according to the image to be displayed.

RGB sensor. The visual sensor used is a Himax HM01B0⁴ Ultra-Low-Power camera with a custom-designed breakout board. The sensor offers an I2C interface for the configuration of the camera settings. Data transmission, instead, is implemented with an 8-bit parallel communication. The camera has an active area of 324x324 pixels, each of which with a side dimension of 3.6 μ m. It can be configured to provide QVGA (320x240) or QQVGA (160x120) images. Unlike the Lepton camera, the chosen Himax module only works in grayscale.

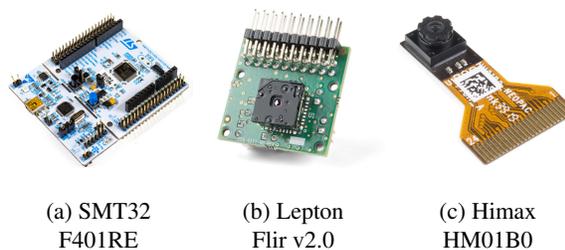


Figure 2: Hardware components used for developing the portable thermal scanner prototype.

Table 2: Camera modules configurations

Camera	Settings	Resolution [p]	Colour
Lepton Flir	RAW 14	80x60	B&W
Lepton Flir	RGB 888	80x60	Colour
Himax	QVGA	320x240	B&W
Himax	QQVG	160x120	B&W

Table 2 summarizes the different tested configurations for the two camera modules. For the final fused image, among the available combinations between IR and grayscale camera, we tested features the Lepton on RGB888 profile and the Himax at QVGA resolution, implementing the appropriate routines to ensure the images can be superimposed.

¹<https://www.st.com/en/microcontrollers-microprocessors/stm32f4-series.html>

²<https://www.st.com/en/evaluation-tools/nucleo-f401re.html>

³<https://www.flir.it/products/lepton/>

⁴<https://www.himax.com.tw/products/cmos-image-sensor/always-on-vision-sensors/hm01b0/>

3.1. Image manipulation

We tested several image processing routines. As previously explained, the images coming from the Himax camera was already of an acceptable quality. Thus, only the thermal image required some enhancements. The most straightforward routines to improve the image quality, feature a custom 3×3 convolution matrix. In our tests we checked the outcomes of several convolutional Gaussian-like filters. However, the image acquired by the camera still resulted quite noisy, driving us towards the adoption of a 3×3 median filter, which is slightly more invasive, compared to a low-pass one, yet better suiting our purpose.

It is well known that median filtering is generally used when dealing with non-uniform noises, which is our case when looking at the Lepton Flir camera. Unlike the averaging method, the noise patches are often partially replaced rather than blurred, thus yielding a cleaner image. The effect of the median filter, although very beneficial for the noisy areas of the image, is often too invasive for the remaining portions of the frame. To balance this, we performed, in addition to the median filter, an averaging step based on the original value of the pixel:

$$p_{i,j} = \alpha \cdot p_{i,j} + (1 - \alpha) \cdot M_F(p_{i,j})$$

where $p_{i,j}$ is the pixel in the i -th row and j -th column of the frame, α is a experimentally found coefficient ($\alpha \in [0, 1]$) and $M_F(p_{i,j})$ is the result of the median filter applied to the pixel $p_{i,j}$. After some experimental tests, we concluded that a suitable value for the α coefficient was 0.6. The values has been chosen manually, by checking the visual consistency of the data.

3.2. Image Fusion

For the thermal scanner image, namely the combination between the IR and the grayscale image, we tested a few options in order to obtain a reliable representation. This has been achieved aligning the images gathered by the two cameras and performing a weighted average between the two, in particular:

$$p_{i,j_F} = \alpha \cdot p_{i,j_T} + (1 - \alpha) \cdot p_{i,j_V}$$

where $p_{i,j_F}, p_{i,j_T}, p_{i,j_V}$ represent the pixel $p_{i,j}$ of the fused image, thermal and grayscale image, respectively, with α , chosen experimentally, set to 0.3. Also in this case, the value has been set manually, especially considering that it only serves to visualization purposes, not affecting the processing pipeline whatsoever.

This proved to be a sufficiently accurate, and turned out to be a valuable qualitative way to check the correctness of the superimposition of the two images. In fact, considering on the one hand the low resolution of the cameras, and, on the other hand, the fact the IR image is positioned at the

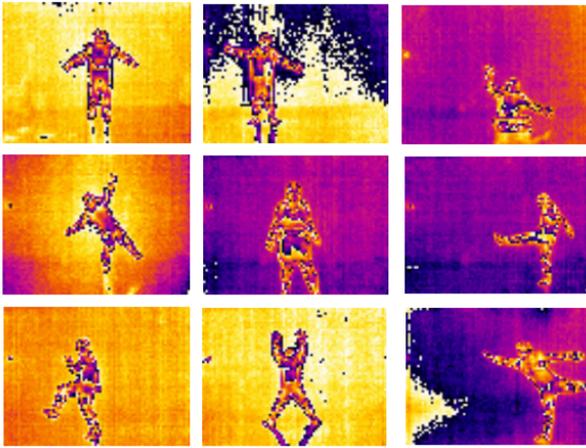
Subject	Male		Female	
	N of images	%	N of images	%
1	750	19.8	74	1.9
2	375	9.9	181	4.8
3	107	2.9	166	4.4
4	88	2.4	105	2.8
5	112	2.9	-	-
6	370	9.8	-	-
7	209	5.5	-	-
8	651	17.3	-	-
9	587	15.6	-	-
Tot	3249	86	526	14

Table 3: Description of dataset, reporting the number of images, and the corresponding percentage in the dataset, for each subject involved.

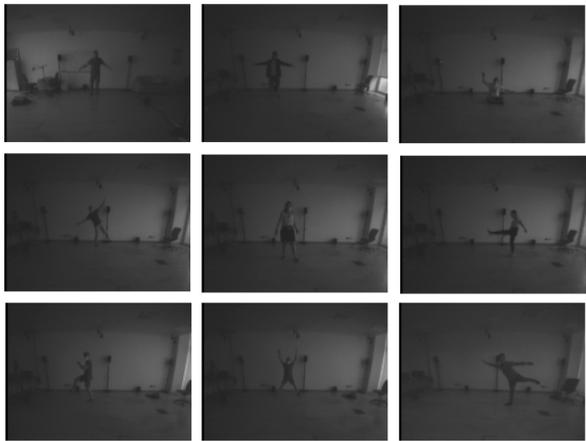
center of the grayscale one, the mismatches due to the differences in the cameras parameters can be considered negligible. It is also to be noted that the spatial resolution and the intrinsic parameters of the two cameras are different, making it impossible to cover the same portion of the observed environment. This would have been possible only by placing the two cameras considerably far apart. However, this would have had two shortcomings: in the first place the two sensors should be positioned at least 30cm apart, with the Himax in front of the Flir IR camera, a quite sub-optimal setup. On the other hand it would have required doubling the size of the IR image. The ultimate solution we picked, is to preserve the original resolutions of the images and fuse the IR image in the central portion of the grayscale one. A few examples are reported in the experimental section.

4. Dataset

To generate a heterogeneous dataset, we set up a shooting session, in which participants have been recorded in different poses. The whole dataset is composed by 3765 gray scale pictures with a resolution of 160×120 pixels. For each picture a corresponding 80×60 pixels IR image was also acquired. This lead to a grand total of 7530 pictures. As shown in Table 3 the users' panel that contributed to the creation of the dataset consists of 13 candidates, 9 male and 4 female. All images were taken in a controlled environment, in which the subjects were free to move. As can be noted from Figure 3, which shows a few samples, the dataset features a wide variety of poses, as well as a reasonably heterogeneous pool of people in terms of body size. It is also worth mentioning that the pictures we have collected do not raise privacy concerns, as the involved subjects are hardly recognizable, especially when considering the IR data.



(a) Lepton Flir camera images



(b) Himax camera images

Figure 3: Sample dataset images.

5. Detection algorithm

Once the correctness of the two images superimposition is verified, we proceeded developing the detection algorithm. We decided to rely on the well-known and established real-time detector YOLO. YOLO is considered among the best-performing state-of-the-art tool for object detection, backed by an ever-growing online community. Among the different versions of YOLO, we had to find out the one ensuring a reliable detection feeding as input the low-resolution images at our disposal.

In order to verify the feasibility of the approach, and remembering that the grayscale image is mostly used as a piece of information supporting the IR data, we decided to train two separate networks, one for each camera, to eventually compare the achieved results.

To do so, an initial dataset consisting of almost 800 images was manually created and labelled, both for the visual

and IR domain. Much of our effort was put into creating a dataset that would be as diverse as possible in terms of the subjects' pose, and their position inside the frame. Afterwards, both of the datasets have been digitally augmented with automatic online tools, thus increasing the size of the dataset by performing common image manipulations such as shear, rotation, vertical flips and colour balancing.

The training phase has been performed on Google Colab. This allowed us to easily share datasets and final results and have access to free external computational resources.

For each training session we have considered a different YOLO version, starting from the recommended configuration file and set of pre-trained weights, and then tuning them based on our specific needs. In the end, we trained 8 different networks (4 for each of the two camera modules), testing the following versions of the network:

- YOLO v3 [46]
- YOLO v3-tiny-3l
- YOLO v4 [47]
- YOLO v4-tiny-3l

Specifically, we wanted to compare the different YOLO versions on two levels: the v3s and v4s families and their corresponding tiny versions. The idea is pretty straightforward; we run the detection algorithm on a benchmark video recorded with our camera modules. Then, we compare the performance of the different YOLO nets in terms of number of detections and average computational time for each detection. The experimental results will be showed and better discussed in 6.

As we intuitively predicted, the network trained with the grayscale camera images tends to yield better and more reliable detections with respect to the IR network, and we assume this is due to the use of data in the visible spectrum, which makes it very comparable to the data originally used to train YOLO. The performance on the IR data are instead lower, and are likely to be bound to the low-resolution images and low robustness with respect to noise sources. With this consideration in mind, our goal to develop an IR dataset by exploiting the more reliable detection offered by the grayscale camera, seems even more appropriate and meaningful.

For the dataset creation, we then relied on the detection accuracy in the grayscale domain, to automatically label the IR data, thanks to the verified quality of the superimposition. The script is meant to work offline, thus diminishing the computational burden of the real-time image-acquisition phase. The overall structure of the code and procedure to be followed is :

1. Visual and IR images are acquired and stored in a memory directory

2. The detection algorithm is run offline on the grayscale images
3. Whenever a detection is returned, the bounding box is properly translated to the correspondent IR image
4. The position and dimension of the bounding box are stored in a label file, associated with the IR image

By iteratively performing such matching operation across all the stored images, we can rely on an automatic labeling tool, that requires minimum supervision to correct the position of a few erroneous detections, thus generating ready-to-use datasets.

6. Experimental results

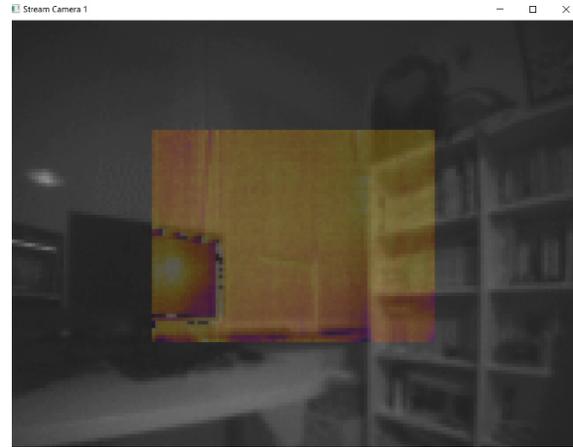
Below, we show, compare and discuss some experimental results. In particular, we can consider the effects of the different post-processing routines that we described in subsection 3.1, compare the detection algorithms among the different YOLO versions and finally evaluate the robustness of the automatically generated IR dataset. For the filtering operations, we report hereafter three different examples (see Figure 4) where the following convolution masks have been used:

$$W_{A_1} = \frac{1}{10} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 6 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad W_{A_2} = \frac{1}{1000} \begin{bmatrix} 25 & 75 & 25 \\ 75 & 600 & 75 \\ 25 & 75 & 25 \end{bmatrix}$$

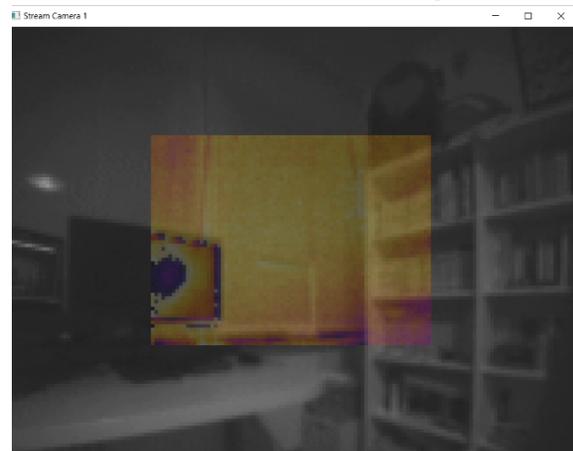
As we can see from the images provided in Figure 4, the difference among the two averaging filters is subtle, while the difference between the effect of the median and the averaging filters is more evident. In particular, it could be pointed out that the convolution matrix W_{A_2} yields, as it was easy to expect, a more rounded blur. The benefits of applying the custom median filter, instead, is very noticeable from the noise point of view. As anticipated, in fact, some of the noisier portions of the frame are considerably enhanced, benefiting the whole image. Therefore, we can safely say that the custom median filter was the best image processing routine among the ones we tested.

When considering one of the project's core objectives, namely the superimposition between the thermal and grayscale camera, the results are indeed satisfying. The support device proved itself to be very functional and easy to use. As a result, the overlapping of the two images is very acceptable, especially in the central portion of the frame. Here, the distortions caused by the different lenses of the camera modules, are hardly noticeable.

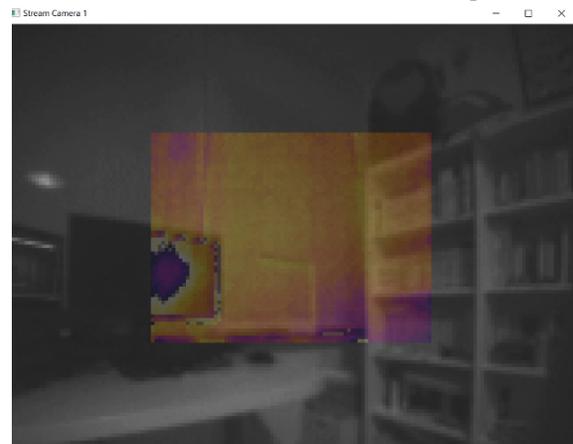
For what concerns the software point of view of the project, all the developed material can be found in a GitHub repository [48]. Here, you will also find a reference to some video footage that we captured with the device. Although



(a) Averaging convolution matrix W_{A_1} (cross)



(b) Averaging convolution matrix W_{A_2}



(c) Custom median filter

Figure 4: Image sample of different superposition approach.

the cameras' performance is more than capable of providing video footage with a satisfying frame rate, this could not be achieved in our particular setup. We found our bottleneck

to be the transmission of the data from the microcontroller to the PC, via the UART port. The low speed of the communication, in fact, dramatically decreased the perceived performance of the cameras.

Although it may not be considered reasonable for many applications, the frame rate is still acceptable for an early stage of the product. Moreover, it should be noted that the frame rate is highly dependent on the amount of data to be transferred. This means, for example, that an RGB footage will inevitably perform worse than a grayscale one due to the presence of three times the amount of data. The same concept also applies to the image's resolution: the lower the amount of data, the faster the frame rate. When combining the two, it is easy to understand why the frame rate of the grayscale IR image is substantially higher than the fused image between RGB IR and grayscale. After the training phase, we move to test the performance of the different network presented. Figures 5 and 6 show the achieved performance, respectively for the detection of human subjects from gray scale images and from IR images. During this evaluation phase, 261 samples were used, among which 243 were the positive (i.e., with human subjects) ones. We performed the same test also with the GPU support enabled, and naturally we saw a noticeable drop in computational time. Figure 7 shows an example of a pair of grey-scale/IR pictures after their evaluation using a YOLO detector.

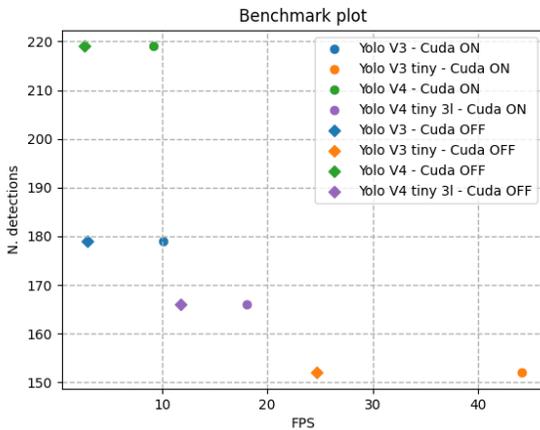


Figure 5: Performance of different Yolo models - Himax

As it was to be expected, the YOLO v4 network ranks first as far as number of detections for both camera modules, closely followed by the YOLO v3. These networks, although allowing to achieve a greater number of detections, are typically much more expensive in terms of computational burden. This will translate, inevitably, in a longer period of time for the detection to be performed, or smaller perceived frames per second (FPS) as showed in the plot. An opposite case can be made for the tiny-3l versions, in

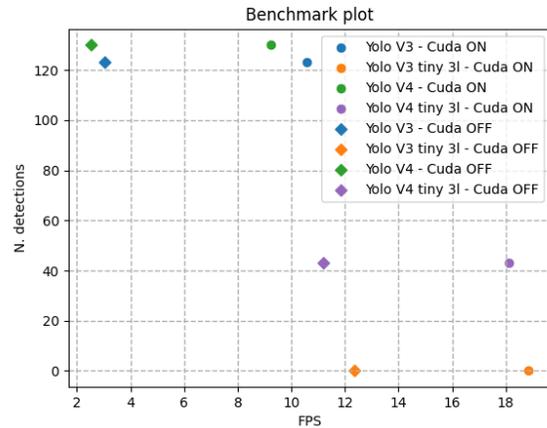
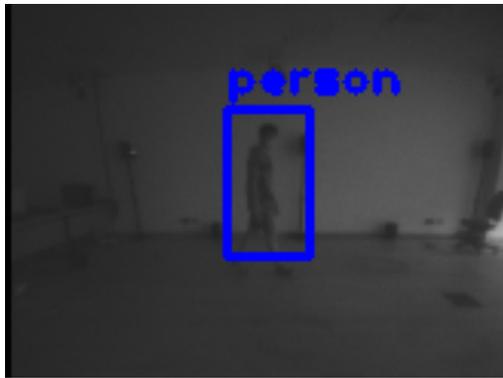


Figure 6: Performance of different Yolo models - Lepton

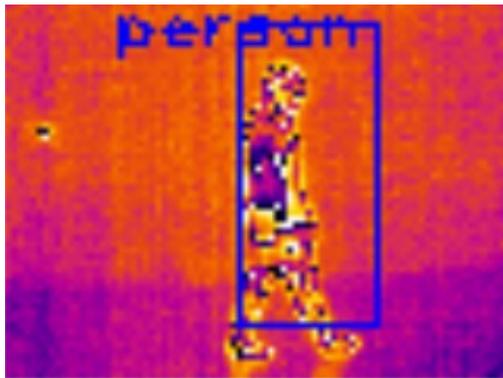
which the focus is shifted more on the computational speed. It is evident how, in general, it exists an underlying compromise between accuracy and computational speed, which will be ruled mainly by the field of application. In this case, since our algorithm for the dataset generation is meant to be run offline, there is really no point in sacrificing accuracy in favour of computational effort. Furthermore, given that our final product, namely the automatically labelled dataset, would be the foundation for future detection algorithms, the possibility of having poor detections was not an option.

When comparing the two plots, a noteworthy aspect of the YOLO v4-tiny-3l shall be mentioned. This network achieved quite different results in terms of number of detections between the two sets of images. For the visual camera, in fact, this version could be a very valuable compromise and could be considered as the network of choice for some real-time applications. Regarding the performance of the other tiny-3l versions, the amount of detections is so small that we discourage their use for this specific application. These performance are most probably due to low resolutions of the images and small scale of the subjects. The performance related to the processing speed refer to the offline execution of the detector. In fact, the main bottleneck is delivering the information from the board to the processing unit, which allows for about 3fps.

On the positive side, an element to be considered is the fact that none of the trained network tend to perform noticeable misdetections.



(a) Black and white frame with YOLO labeling



(b) IR frame with automatic labeling

Figure 7: Dataset sample image after the evaluation using a YOLO network. In (a) a detected human from a grey scale image while in (b) the same subject detected from the IR image.

7. Conclusion

We presented a portable, low-cost thermal scanner prototype consisting of a thermal sensor assisted by a gray-scale camera. The device is built with COTS components and is connected to a separate computer for the analysis. The system has been tested in the context of a people detection problem, which we found to be an appropriate use-case to set up the learning framework together with the corresponding IR dataset, collected and annotated via the paired gray-scale images. Results show that the use of IR data can be beneficial, also in light of a more privacy-preserving detection. At the current stage we believe that better accuracy could be reached by further expanding the IR dataset and introducing a filtering stage capable of better highlighting the shapes of the moving subjects. Future works will encompass both the expansion of the dataset in order to increment the detector performance and the development of a more compact, battery powered and standalone thermal scanner in order to ease the creation of low-resolution IR datasets for people detection.

References

- [1] D. T. Nguyen, W. Li, and P. O. Ogunbona, "Human detection from images and videos: A survey," *Pattern Recognition*, vol. 51, pp. 148–175, 2016. 1
- [2] D. Liciotti, M. Paolanti, E. Frontoni, and P. Zingaretti, "People detection and tracking from an rgb-d camera in top-view configuration: Review of challenges and applications," in *New Trends in Image Analysis and Processing – ICIAP 2017* (S. Battiato, G. M. Farinella, M. Leo, and G. Gallo, eds.), (Cham), pp. 207–218, Springer International Publishing, 2017. 1
- [3] S. K. Sharma, R. Agrawal, S. Srivastava, and D. K. Singh, "Review of human detection techniques in night vision," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2216–2220, 2017. 1
- [4] O. Ukimura, *Image fusion*. BoD–Books on Demand, 2011. 1
- [5] J. Davis and V. Sharma, "Robust detection of people in thermal imagery," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 4, pp. 713–716 Vol.4, 2004. 1
- [6] M. Rai, T. Maity, and R. Yadav, "Thermal imaging system and its real time applications: a survey," *Journal of Engineering Technology*, vol. 6, no. 2, pp. 290–303, 2017. 1, 2
- [7] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Machine Vision and Applications*, vol. 25, pp. 245–262, Jan 2014. 1, 2
- [8] G. Messina and G. Modica, "Applications of uav thermal imagery in precision agriculture: State of the art and future research outlook," *Remote Sensing*, vol. 12, no. 9, 2020. 1, 2
- [9] Y. Jeong, K. Yoon, and K. Joung, "Probabilistic method to determine human subjects for low-resolution thermal imaging sensor," in *2014 IEEE Sensors Applications Symposium (SAS)*, pp. 97–102, 2014. 1
- [10] G. Cerutti, B. Milosevic, and E. Farella, "Outdoor people detection in low resolution thermal images," in *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pp. 1–6, 2018. 1, 2
- [11] G. Cerutti, R. Prasad, and E. Farella, "Convolutional neural network on embedded platform for people presence detection in low resolution thermal images," in *ICASSP 2019 - 2019 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7610–7614, 2019. 1, 2

- [12] P. Talluri and M. Dua, “Low-resolution human identification in thermal imagery,” in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1283–1287, 2020. 1, 2
- [13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” 2016. 2
- [14] R. Vadivambal and D. S. Jayas, “Applications of thermal imaging in agriculture and food industry—a review,” *Food and Bioprocess Technology*, vol. 4, pp. 186–199, Feb 2011. 2
- [15] R. Ishimwe, K. Abutaleb, F. Ahmed, *et al.*, “Applications of thermal imaging in agriculture—a review,” *Advances in remote Sensing*, vol. 3, no. 03, p. 128, 2014. 2
- [16] V. Santhi, “Recent advances in applied thermal imaging for industrial applications,” 2017. 2
- [17] E. Ring and K. Ammer, “Infrared thermal imaging in medicine,” *Physiological measurement*, vol. 33, no. 3, p. R33, 2012. 2
- [18] S. Khanal, J. Fulton, and S. Shearer, “An overview of current and potential applications of thermal remote sensing in precision agriculture,” *Computers and Electronics in Agriculture*, vol. 139, pp. 22–32, 2017. 2
- [19] J. A. Tsanakas, L. D. Ha, and F. Al Shakarchi, “Advanced inspection of photovoltaic installations by aerial triangulation and terrestrial georeferencing of thermal/visual imagery,” *Renewable Energy*, vol. 102, pp. 224–233, 2017. 2
- [20] Y. Zefri, A. Elkctani, I. Sebari, and S. A. Lamallam, “Inspection of photovoltaic installations by thermovisual uav imagery application case: Morocco,” in *2017 International Renewable and Sustainable Energy Conference (IRSEC)*, pp. 1–6, IEEE, 2017. 2
- [21] I. Shopovska, L. Jovanov, and W. Philips, “Deep visible and thermal image fusion for enhanced pedestrian visibility,” *Sensors*, vol. 19, no. 17, p. 3727, 2019. 2
- [22] V. John, S. Mita, Z. Liu, and B. Qi, “Pedestrian detection in thermal images using adaptive fuzzy c-means clustering and convolutional neural networks,” in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, pp. 246–249, 2015. 2
- [23] C. Lile and L. Yiqun, “Anomaly detection in thermal images using deep neural networks,” in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2299–2303, 2017. 2
- [24] Y. Li, J. X. Gu, D. Zhen, M. Xu, and A. Ball, “An evaluation of gearbox condition monitoring using infrared thermal images applied with convolutional neural networks,” *Sensors*, vol. 19, no. 9, 2019. 2
- [25] C. D. Rodin, L. N. de Lima, F. A. de Alcantara Andrade, D. B. Haddad, T. A. Johansen, and R. Storvold, “Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2018. 2
- [26] M. Marzec, A. Lamża, Z. Wróbel, and A. Dziech, “Fast eye localization from thermal images using neural networks,” *Multimedia Tools and Applications*, Nov 2016. 2
- [27] S. Farokhi, J. Flusser, and U. Ullah Sheikh, “Near infrared face recognition: A literature survey,” *Computer Science Review*, vol. 21, pp. 1–17, 2016. 2
- [28] S. Coşar, Z. Yan, F. Zhao, T. Lambrou, S. Yue, and N. Bellotto, “Thermal camera based physiological monitoring with an assistive robot,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5010–5013, 2018. 3
- [29] V. Espinosa-Duró, M. Faundez-Zanuy, and J. Mekyska, “A new face database simultaneously acquired in visible, near-infrared and thermal spectrums,” *Cognitive Computation*, vol. 5, no. 1, pp. 119–135, 2013. 3
- [30] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon, “Multispectral pedestrian detection: Benchmark dataset and baseline,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1037–1045, 2015. 3
- [31] T. Kawashima, Y. Kawanishi, I. Ide, H. Murase, D. Deguchi, T. Aizawa, and M. Kawade, “Action recognition from extremely low-resolution thermal image sequence,” in *Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 1–6, Aug. 2017. 3
- [32] Q. Liu, X. Li, Z. He, C. Li, J. Li, Z. Zhou, D. Yuan, J. Li, K. Yang, N. Fan, *et al.*, “Lsotb-tir: A large-scale

- high-diversity thermal infrared object tracking benchmark,” in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 3847–3856, 2020. 3
- [33] M. Krišto and M. Ivašić-Kos, “Thermal imaging dataset for person detection,” in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1126–1131, 2019. 3
- [34] “Flir thermal dataset for algorithm training.” <https://www.flir.com/oem/adas/adas-dataset-form/>. 3
- [35] D. Tan, K. Huang, S. Yu, and T. Tan, “Efficient night gait recognition based on template matching,” in *18th International Conference on Pattern Recognition (ICPR’06)*, vol. 3, pp. 1000–1003, 2006. 3
- [36] M. Teutsch, T. Muller, M. Huber, and J. Beyerer, “Low resolution person detection with a moving thermal infrared camera by hot spot classification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 209–216, 2014. 2
- [37] L. Grewe, S. Choudhary, E. Gallegos, D. P. Jain, and P. Aguilera, “Low-resolution infrared temperature analysis for disease situation awareness via machine learning on a mobile platform,” in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXX* (I. Kadar, E. P. Blasch, and L. L. Grewe, eds.), vol. 11756, pp. 287 – 299, International Society for Optics and Photonics, SPIE, 2021. 2
- [38] L. Tao, T. Volonakis, B. Tan, Y. Jing, K. Chetty, and M. Smith, “Home activity monitoring using low resolution infrared sensor,” 2018. 2
- [39] C. Herrmann, T. Müller, D. Willersinn, and J. Beyerer, “Real-time person detection in low-resolution thermal infrared imagery with MSER and CNNs,” in *Electro-Optical and Infrared Systems: Technology and Applications XIII* (D. A. Huckridge, R. Ebert, and S. T. Lee, eds.), vol. 9987, pp. 166 – 173, International Society for Optics and Photonics, SPIE, 2016. 2
- [40] Y. Ma, X. Wu, G. Yu, Y. Xu, and Y. Wang, “Pedestrian detection and tracking from low-resolution unmanned aerial vehicle thermal imagery,” *Sensors*, vol. 16, no. 4, 2016. 2
- [41] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, J. Hai, and K. He, “A survey of infrared and visual image fusion methods,” *Infrared Physics & Technology*, vol. 85, pp. 478–501, 2017. 2
- [42] J. Ma, Y. Ma, and C. Li, “Infrared and visible image fusion methods and applications: A survey,” *Information Fusion*, vol. 45, pp. 153–178, 2019. 2
- [43] B. Meher, S. Agrawal, R. Panda, and A. Abraham, “A survey on region based image fusion methods,” *Information Fusion*, vol. 48, pp. 119–132, 2019. 2
- [44] H. Kaur, D. Koundal, and V. Kadyan, “Image fusion techniques: A survey,” *Archives of Computational Methods in Engineering*, Jan 2021. 2
- [45] T. Sosnowski, G. Bieszczad, H. Madura, and M. Kastek, “Thermovision system for flying objects detection,” in *2018 Baltic URSI Symposium (URSI)*, pp. 141–144, IEEE, 2018. 2
- [46] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 2018. 5
- [47] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” 2020. 5
- [48] “Github repository for the dataset.” https://github.com/AlessandroAvi/Thermocamera_Lepton_Himax. 6