DOI 10.5162/SMSI2025/D1.4

# Development of a Computationally Efficient Person Detection Procedure for Low-Resolution Infrared Sensor Arrays

*Alexander Rehmer*
Heimann Sensor GmbH, Maria-Reiche-Str. 1, 01109 Dresden, Germany
rehmer@heimannsensor.com

**Summary:** A computationally highly efficient Computer Vision algorithm for detecting and tracking persons in low-resolution infrared images is introduced an its performance compared to a Deep Learning model in a case study.

**Keywords:** computer vision, object detection, person detection, infrared, thermopiles

## Motivation

The progress made in Computer Vision during over the last two decades can mainly be attributed to the development and application of very complex models, which require virtually unlimited amounts of data as well as the employment of massive resources (storage, egenery, time, money) . Applicatons which pose restrictions on the availability of any of these resources (e.g. low-cost solutions, edge-computing applications, application on data from non-ubiquotus sensors) can not benefit from the progress made. Therefore, for some applications the development of highly tailored Computer Vision solutions is still necessary. Such a solution is presented in this work.

## Introduction

Object Detection is one of the main tasks in computer vision. Its comprises the correct classification and localization of an instance of a semantic class in a digital image or video [1]. Localization is typically performed by drawing a bounding box as close as possible around the detected instance. Object recognition, as opposed to object detection, merely detects instances of a class, but does not necessarily provide their location. Object tracking is the task of tracking an identified instance of a class as it moves through the image plane from frame to frame. Since this entails an initial identification of an instance and some sort of re-identifaction in subsequent frames, object tracking usually presupposes object detection. Object detection algorithms, henceforth referred to as detectors, can be roughly divided into two classes [2]:

- Two-Stage Detectors work in two consecutive stages: A region proposal stage and a classification stage. During the first stage an algorithm identifies regions of interest (RoI) in an image that may contain an object and in the second stage these proposed regions are evaluated by one or multiple classifiers on whether they contain an instance of the
- One-Stage Detectors process the whole image in a single step by dividing it into a grid of cells and trying to detect objects centered in each of the cells.

In theory, One-Stage Detectors should provide more accurate results than Two-Stage Detectors, since they can be trained end-to-end. This is not possible with Two-Stage Detectors, since proposal generator and classifier are two seperate models. In practive however, Two-Stage Detectors surprisingly outperform One-Stage Detectors [3, 4]. While Two-Stage detectors can be realized using computationally very efficient traditional Computer Vision methods, One-Stage detectors are exclusively Deep Learning models. As such they need large amounts of training data (millions of datapoints) to achieve satisfactory performance as well as powerful hardware to process a frame in a reasonable amount of time [2].

To give even a brief overview over the current state of the art of tracking algorithms would exceed the scope of this article. Nevertheless it should at least be mentioned, that a detector can be extended of even be converted to become a *Tracktor* [5], i.e. a detector with tracking capabilities.

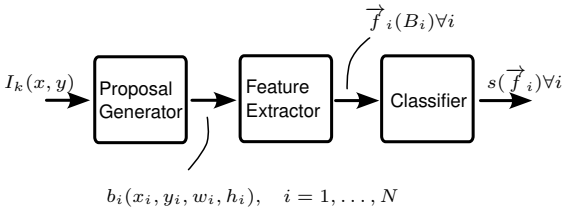


*Fig. 1: Processing steps of a Two-Stage Detector.*

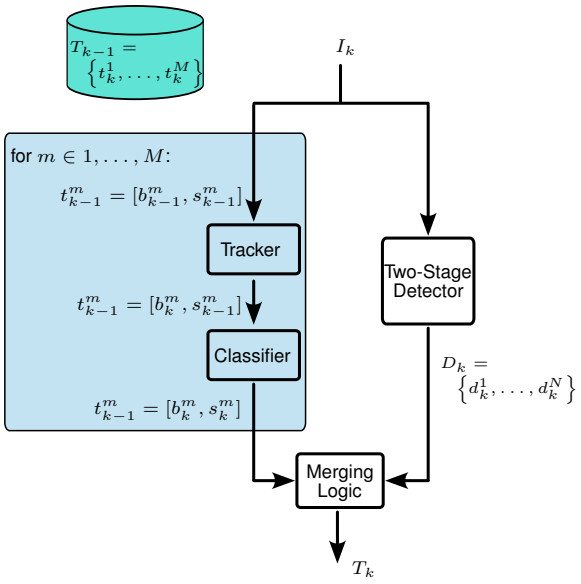


*Fig. 2: Processing steps of the LRI-Tracktor.*

## LRI-Tracktor: A Person Detection and Tracking Procedure for Low-Resolution Infrared Sensor Arrays

Due to the higher accuracy and the possibility to employ computationally efficient traditional Computer Vision methods it was decided to developed a Two-Stage Detector specifically tailored to low-resolution infrared images. To this end, for each of the processing steps in Fig. 1 a method was carefully chosen (or developed if necessary) that fit the characteristics of the sensor data best and therefore yielded the highest performance.

In order to further enhance the performance of the developed detector, i.e. to handle occlusion, maintain instance identities across frames and improve detection accuracy, it was extended with tracking mechanisms. The resulting algorithm is named LRI-Tracktor due to the circumstance that it is tailored to **L**ow-**R**esolution **I**nfrared images. Fig. 2 provides a rough insight into the inner workings of the LRI-Tracktor. It maintains a list $T_k$ of $M$ active tracks $\boldsymbol{t}_k^m$, $m = 1, \ldots, M$, i.e. objects to be tracked. Each track consists of bounding box coordinates $\boldsymbol{b}_k^m$ and a classifier score $s_k^m$, i.e. $\boldsymbol{t}_k^m = [\boldsymbol{b}_k^m, s_k^m]$. The LRI-Tracktor first applies a tracking procedure to each of the tracks, shifting their bounding box positions from frame $I_{k-1}$ to new positions in the new frame $I_k$, i.e. $\boldsymbol{b}_{k-1}^m \rightarrow \boldsymbol{b}_k^m$. Each of the shifted boxes is then evaluated by the detector's classifier, yielding a new score for each track, i.e. $s_{k-1}^m \rightarrow s_k^m$. The frame $I_k$ is passed to the detector as well, yielding a number of detections $D_k$. Finally, the updated tracks $\hat{T}_k$ and the detections $D_k$ are merged to the final result $T_k$ using simple logic, e.g. killing tracks with low scores, initializing new tracks and updating tracks that have been found by the detector.

*(a) 32x32* *(b) 60x40*

*Fig. 3: Frames containing one person from a 32x32 and a 60x40 pixel thermopile array.*

*Tab. 1: Performance of YOLOv5m and the presented LRI-Tracktor on test data. Performance was measured in terms of the Mean Average Precision ($\mathrm{mAP}$).*

| | | Num. of Persons | | |
|---|---|---|---|---|
| | | 3 | 4 | 5 |
| YOLOv5m | $\mathrm{mAP}_{.5}$ | **0.84** | **0.80** | **0.55** |
| | $\mathrm{mAP}_{>.5}$ | 0.36 | 0.28 | 0.15 |
| LRI-Tracktor | $\mathrm{mAP}_{.5}$ | 0.76 | 0.64 | 0.40 |
| | $\mathrm{mAP}_{>.5}$ | **0.36** | **0.32** | **0.16** |

## Case study

The purpose of the following case study is to show that the developed LRI-Tracktor can achieve similar performance to state-of-the-art Deep Learning Computer Vision models on low-resolution infrared image sequence. As comparison model to the LRI-Detector the YOLOv5m Single-Stage detector was chosen. Unfortunately, the data on which YOLOv5m was evaluated is not available anymore, therefore the performance metric are not directly comparable. The LRI-Detector was applied to data of a 60x40 pixels thermopile array (perspective: bird's-eye view). YOLOv5m was applied to data of a 32x32 pixels thermopile array (perspective: top view). Fig. 3 shows an exemplary frame of each thermopile array. Multiple datasets were recorded with each sensor, containing from three up two five persons. The datasets were split into training and validation datasets. The performance on the validation data is documented in Tbl. 1. The results show, that YOLOv5m detects persons correctly more often than the LRI-Tracktor, but locates them only with low precision (rows with $\mathrm{mAP}_{.5}$). THe LRI-Tracktor on the other hand correctly detects persons with higher precision more often (rows with $\mathrm{mAP}_{>.5}$).

## References

[1] A. Mauri, R. Khemmar, B. Decoux, N. Ragot, R. Rossi, R. Trabelsi, R. Boutteau, J.-Y. Ertaud, and X. Savatier, "Deep learning for real-time 3d multi-object detection, localisation, and tracking: Application to smart mobility," *Sensors*, vol. 20, no. 2, p. 532, 2020.

[2] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference*

*(CVC), Volume 1 1*, pp. 128–144, Springer, 2020.

[3] X. Lu, Q. Li, B. Li, and J. Yan, "Mimicdet: Bridging the gap between one-stage and two-stage object detection," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, pp. 541–557, Springer, 2020.

[4] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, and J. García-Gutiérrez, "On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data," *Remote Sensing*, vol. 13, no. 1, p. 89, 2020.

[5] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 941–951, 2019.