Poster: TheDet: A Machine Learning-based Privacy-preserving Occupancy Estimation Method

Shuai Zhu, Joakim Eriksson
RISE Research Institutes of Sweden
shuai.zhu@ri.se
joakim.eriksson@ri.se

Thiemo Voigt
RISE Research Institutes of Sweden and Uppsala University
thiemo.voigt@it.uu.se

Daniel F. Perez-Ramirez
RISE Research Institutes of Sweden and KTH Royal Institute of Technology
daniel.perez@ri.se

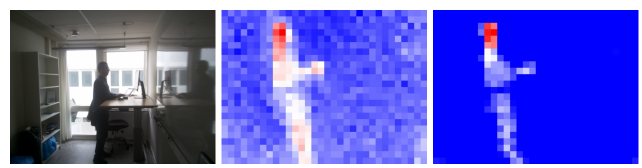
1 Motivation

Energy provision for buildings accounts for 40% of the total energy consumption in the EU, according to the European Union 2020 Energy Policy Review [4]. Therefore, indoor occupancy estimation has attracted much attention in recent years as an important component of indoor energy management systems. However, existing methods commonly suffer from some limitations. For example, methods use conventional cameras [3] which is generally not privacy-preserving or multiple environmental sensors [2] which increases cost and maintenance efforts. On the other hand, multiple object detection (MOD) and multiple object tracking (MOT) have been extensively studied in recent years as important topics in the computer vision field [7]. In this paper, we present TheDet, a privacy-preserving method that uses a low-resolution thermal camera together with MOD and MOT techniques to perform occupancy estimation.

2 Design of TheDet

Different from our earlier work [8], we design TheDet based on a more scalable and lightweight neural network, which enables us to easily train models with suitable sizes for various IoT devices. With the same training dataset [9], in contrast to earlier work, TheDet performs much better with a smaller model in detection tasks. For tracking tasks, TheDet reaches the same level of performance facing a more challenging evaluation, where the testing data contains diverse target places and multiple people, rather than one place and one person.

Hardware. We attach an MLX90640 to a Raspberry Pi device. The MLX90640 is a low-resolution thermal camera with 32×24 pixels. The camera is able to capture infrared radiation emitted from warm objects, and only outputs the temperature distribution of its field of view (FoV). Figure 1a and Figure 1b respectively show a normal image and a low-resolution thermal image. By contrast, the thermal image does not contain any private information and filters most of the irrelevant objects for the occupancy estimation task, such as furniture. TheDet utilizes this feature of the low-resolution thermal image to cope with privacy issues, such as leaking information about people or personal belongings.



(a) Regular image. (b) Thermal image. (c) Clear thermal image.

Figure 1: (a) shows an image from a regular camera. (b) and (c) are from the same low-resolution thermal camera. (b) is blurry due to ambient heat. (c) has a pure background because we use a low threshold higher than the lowest environmental temperature to filter ambient heat.

Software. As illustrated in Figure 2, TheDet mainly consists of three components: Data-preprocessing component, Detection-based and Tracking-based estimation methods. The data-preprocessing component maps temperature information to colors. We use a low threshold higher than the lowest environmental temperature to make the system insensitive to the ambient heat so that images have a pure background, as illustrated in Figure 1c. For the high threshold, we directly use the highest environmental temperature. For the design of detection-based and tracking-based estimation methods, there is a trade-off between efficiency and performance since TheDet is designed for IoT devices with limited computing resources. Therefore, we design the detection-based method based on EfficientNet [6], a scalable and efficient neural network designed for IoT devices. The tracking-based method is also based on a lightweight object tracking framework, Simple Online and Realtime Tracking [1].

TheDet uses two working modes, detection mode and tracking mode, to respectively cope with different application scenarios. The detection mode is used to estimate the number of people in scenarios where one camera can cover the whole target space, and the tracking mode is used to cope with scenarios where one camera cannot cover the target space. In the detection mode, TheDet captures images of the

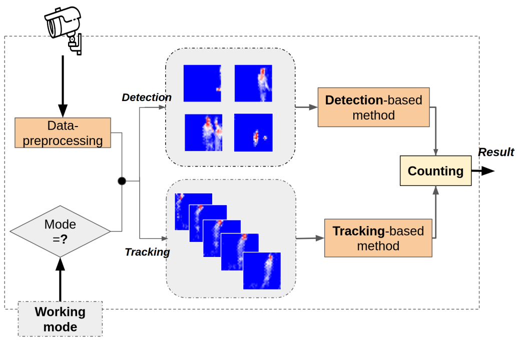


Figure 2: Overview of TheDet. TheDet mainly consists of three components: Data-preprocessing component, Detection-based method, and Tracking-based method. It has two working modes: detection mode and tracking mode.

(a) Input

(b) Output

Figure 3: (a) shows the input image of the detection-based method. (b) visualizes the output. The left object is a person with 98% probability, and the right object is the other thermal object with 96% probability.

Table 1: The detection performance of TheDet and Zhu [8].

Detector	mAP	mAP_{50}	mAP_{75}	AP_{others}	AP_{person}
TheDet	0.807	0.963	0.888	0.711	0.905
Zhu [8]	-	0.568	-	-	-

Table 2: The tracking performance of TheDet and Zhu [8].

Tracker	IDF_1	$MOTA$	$MOTP$	IDP	IDR
TheDet	0.594	0.432	0.719	0.912	0.440
Zhu [8]	0.596	0.473	0.781	-	-

target space and feeds them to the detection-based method as input. Then, the detection-based method outputs locations and categories of thermal objects in images, as shown in Figure 3. Finally, TheDet counts the number of people in the target place. In the tracking mode, TheDet captures sequence image frames of the entrance of the target space and feeds them to the tracking-based method. Then, the tracking-based method tracks the people in frames and saves their last coordinates appearing in the FoV. Finally, TheDet figures out whether tracked people enter or leave the target space. The number of people is maintained in memory and is updated when people enter or leave the room, for example, plus 1 when a person enters.

3 Evaluation

To evaluate TheDet, we create a testing dataset that is collected from two different office rooms for five weekdays. We separately evaluate the performance of the object detection and the tracking methods, as well as the accuracy of occupancy estimation. For the detection and tracking evaluation, we apply the most widely used metrics in these fields [5]. We train a series of detectors of different sizes. All of them perform better than our previous detector [8]. The largest detector is $19.6MB$ which has around $\frac{2}{3}$ of the size of our previous detector ($31.7MB$). Table 1 shows the performance of TheDet with the largest detector and our earlier work. In addition, TheDet obtains the same performance in a more challenging testing tracking scenario, as illustrated in Table 2.

We also evaluate the accuracy of occupancy estimation. In the detection mode, TheDet obtains an accuracy of 98.0% when detecting whether the target room is empty or not and an accuracy of 92.0% when estimating the number of people within the target place. In the tracking mode, TheDet reaches an accuracy of 87.5% when judging whether people enter or leave the target space.

4 Conclusion

We present TheDet, a machine learning-based privacy-preserving occupancy estimation method. TheDet achieves an accuracy of up to 92% for occupancy estimation and 98% accuracy for occupancy detection in scenarios where one camera can cover the target space. In addition, TheDet obtains 87.5% accuracy of tracking and judging whether people enter or leave a room in scenarios where one camera cannot cover the target space.

Acknowledgements. This project is supported by the Swedish Foundation for Strategic Research.

5 References

[1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft. Simple online and realtime tracking. *CoRR*, abs/1602.00763, 2016.

[2] Z. Chen, C. Jiang, and L. Xie. Building occupancy estimation and detection: A review. *Energy and Buildings*, 169:260–270, 2018.

[3] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):267–282, 2008.

[4] IEA 2020. European union 2020, iea, paris. Retrieved 23 May, 2022 from https://www.iea.org/reports/european-union-2020.

[5] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):319–336, 2009.

[6] M. Tan and Q. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[7] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018:7068349, 2018.

[8] S. Zhu. Privacy-preserving building occupancy estimation via low-resolution infrared thermal cameras. Master's thesis, KTH, School of Electrical Engineering and Computer Science (EECS), 2021.

[9] S. Zhu, T. Voigt, D. F. Perez-Ramirez, and J. Eriksson. A low-resolution infrared thermal dataset and potential privacy-preserving applications. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, SenSys '21, page 552–555, New York, NY, USA, 2021. Association for Computing Machinery.