Doors as Visual Landmarks for Indoor Positioning

Miroslav Opiela[1]

[1]Faculty of Science, Institute of Computer Science, Pavol Jozef Šafárik University in Košice, 04001 Košice, Slovakia

Abstract

Smartphone-based indoor positioning using Pedestrian Dead Reckoning (PDR) accumulates error over time due to inaccurate sensor measurements and imperfect step length estimation. The position can be corrected using Bayesian filtering combined with additional sources of information, such as maps or Wi-Fi signals. This paper explores the use of door detection from camera images to correct position estimates. An off-the-shelf deep learning-based method (Grounding DINO) is employed for door detection. Detected doors are compared with the actual visible doors according to the map, and the position is corrected accordingly. An experiment conducted on a 30-meter straight path demonstrates a reduction in positioning error from 3.75 m to 1.15 m. Additionally, the paper discusses the anonymization of individuals captured in camera images to address privacy concerns.

Keywords

indoor positioning, door detection, anonymization

1. Introduction

Indoor positioning is not a unified discipline. It encompasses a wide range of use cases, solutions, technologies, and techniques. Each use case comes with its own unique challenges. In the case of smartphone-based indoor positioning, one major challenge is the freedom of movement of the device, as opposed to the more constrained and predictable motion of robots or fixed-position sensors. Another persistent issue lies in the precision of the smartphone sensors.

Smartphone-based positioning commonly relies on Pedestrian Dead Reckoning (PDR), which is a technique that estimates the user relative position based on sensor measurements. Typically, it computes displacement when a step is detected. However, inaccuracies in sensor readings, incorrect step length estimation, and missed or false step detections result in the accumulation of positioning errors over time.

To mitigate these errors, bayesian filtering is employed in various forms [1]. The most common approaches in indoor positioning are the Kalman filter and particle filter, often adapted in their extended or modified forms to better handle the system's non-linearities. These methods treat the user position as a random variable, representing the system state. Filtering involves two key phases: the transition (prediction) phase, where user movement introduces uncertainty, and the correction (observation) phase, which integrates additional information to reduce that uncertainty. Since PDR estimates relative position, these corrections may incorporate information from external sources such as Wi-Fi or Bluetooth (BLE) signals to improve absolute positioning accuracy. Additionally, map-based models (derived from building floor plans) can be used for correcting position estimates [2] or to constrain movement models directly [3].

Nevertheless, most of these correction methods depend on building infrastructure or require maintenance effort (e.g., maintaining a Wi-Fi fingerprint database). An alternative could be the visual approach, which uses a smartphone's camera as a positioning aid. A camera-based track was included in some IPIN competition editions, e.g., 2018 (on-site) [4], 2019 (on-site) [5], and 2022 (off-site) [6], but it has not been featured regularly. Nonetheless, the rapid advances in artificial intelligence (AI) and computer vision suggest that visual information could play a much larger role in indoor positioning.

IPIN-WCAL 2025: Workshop for Computing & Advanced Localization at the Fifteenth International Conference on Indoor Positioning and Indoor Navigation, September 15–18, 2025, Tampere, Finland

miroslav.opiela@upjs.sk (M. Opiela)

0000-0001-8802-4442 (M. Opiela)

The main challenges for vision-based positioning include accuracy, computational efficiency, and privacy. This paper explores the accuracy of a visual positioning method using off-the-shelf AI models in a novel scenario where doors are used as landmarks. It also addresses the privacy aspect through anonymization techniques that do not significantly affect positioning performance. Direct real-time requirements are not addressed. However, when multiple options exist, faster calculations are preferred. The goal is to evaluate the viability of this approach. Promising results could pave the way for more advanced and privacy-preserving visual positioning solutions.

The paper is structured as follows: Chapter 2 presents related work on indoor positioning, door detection, and anonymization. Chapter 3 introduces the proposed positioning method. Chapter 4 describes the experimental setup and evaluates the performance of the method. Finally, results are discussed and possible directions for future research are outlined.

2. Background and related work

2.1. Indoor positioning framework

Indoor positioning allows for various techniques [7]. The core of the positioning method adopted in this paper is based on author's previous work [2]. Bayes filters are commonly used in positioning systems, as they can handle inaccurate inputs and provide a framework for position estimation.

In smartphone-based positioning, a popular choice is the Particle filter (e.g., [8]), where the state is represented by a set of particles with assigned belief values. Particles are displaced according to detected steps, their weights can be updated based on observations, and then particles are resampled. The model enables focus on areas with the highest probability of the true position, achieving better precision. However, sample impoverishment can occur when particles concentrate in areas of high belief, which may lead to position loss if the assumption is incorrect. Moreover, the Particle filter is a stochastic method that introduces randomness in particle movement.

The Particle filter allows for fast computation and good performance even with a relatively small number of particles, and can represent a more complex system state, including position, orientation, speed, etc. In this paper, which is focused on door detection integration, we use a grid-based filter [9], which is deterministic. The map is tessellated into a regular grid, and belief is computed for all points (i.e., centers of grid cells). The basic grid-based filter preserves all information but lacks the ability to focus on specific areas with higher resolution.

The computation of the posterior belief, based on the prior distribution, is triggered by a detected step. Steps are identified using sensor measurements [10]. For the Bayes filter transition phase, both orientation and distance are needed. Step length can be computed from sensors using step-frequency-based, acceleration-based, angle-based, or multiparameter approaches [11].

2.2. Door detection

Doors can be detected in images using computer vision methods such as edge and corner detection [12] or line detection [13]. These methods require a certain amount of computation, and their accuracy is limited. Moreover, most door detection papers are evaluated in outdoor scenarios, where a wide range of distinct door appearances is present.

For indoor positioning within a specific building, a more specialized method would be preferable. Neural networks can be trained to detect doors in images. A promising approach is to combine existing robust object detection models [14] with fine-tuning tailored to the specific building. Antonazzi et al. [15] propose a door detection method for robots. They perform fine-tuning of a deep neural network using synthetic data generated from photorealistic simulation environments. Zhang et al. [16] improve the original YOLOv3 model by integrating DenseNet blocks in the context of door and window detection for robots. Their solution includes additional techniques that enhance accuracy without significantly increasing computation time.

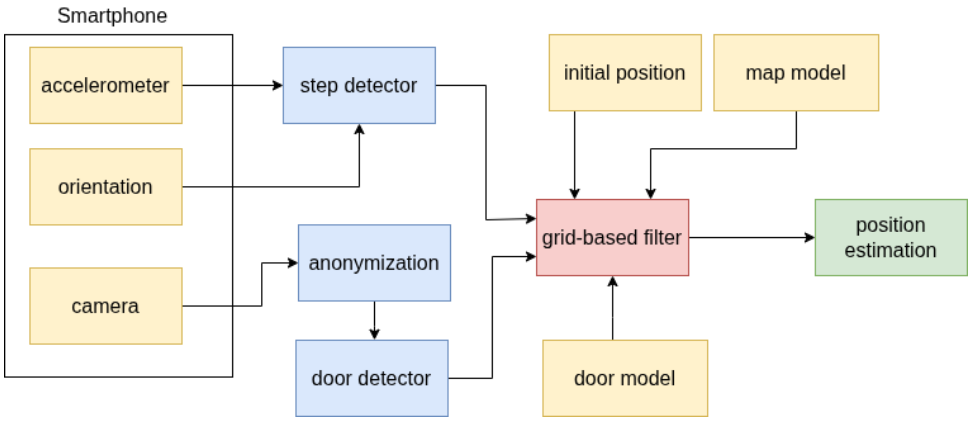

Figure 1: Diagram of the positioning framework.

YOLO models [17] typically require fine-tuning to detect doors, as the COCO dataset [18], on which most YOLO models are pretrained, does not include *door* as a separate category. However, there are object detectors capable of detecting doors without the need for additional training. Grounding DINO [19] enables detection and localization of objects in images using free-text prompts, without relying on predefined class labels. It combines a transformer-based image encoder with a language-guided decoder, allowing flexible object detection queries expressed in natural language.

Although a fine-tuned model would likely achieve higher accuracy, for the purpose of demonstrating how door detection can be integrated into an indoor positioning system, this off-the-shelf approach is sufficient. Moreover, door detection serves not as the core positioning method, but rather as a correction mechanism to adjust PDR estimations.

2.3. Anonymization

To address the privacy aspect of camera-based positioning applications, the primary concern is the identification of individuals captured in the image. Various techniques exist to anonymize people in images. Face detection using Haar features is a commonly used approach [20], as it allows for fast classification of image regions. Detected faces can then be anonymized using techniques such as black boxes, blurring, or pixelation [21].

In recent years, several deep learning-based methods have emerged, offering more advanced capabilities. Deep Privacy 2 [22] uses a Generative Adversarial Network to detect human faces or full bodies and replaces them with synthetic counterparts. LaMa [23] is a mask inpainting method capable of handling complex images using Fourier convolutions. It removes detected individuals from images through inpainting. For the purpose of door detection, the inpainting approach may be more suitable, as it presents a simplified image with fewer distracting objects. However, this method is not yet capable of real-time processing.

3. System overview

Figure 1 presents an overview of the proposed system. The method is composed of the following steps:

- Initialization of the system: the map is tessellated into a regular grid based on floor plans. The initial position is provided, and a door model is created from the map.
- Accelerometer sensor measurements are processed by a step detector, which triggers position estimation (performed by the grid-based filter). A low-pass filter is applied to smooth the data.
- Orientation is obtained from the rotation vector and supplied to the step detector.
- Step detection is performed using a four-phase identification method [9]. Step length is calculated based on the user's height and step frequency using the formula introduced by Villien et al. [24].

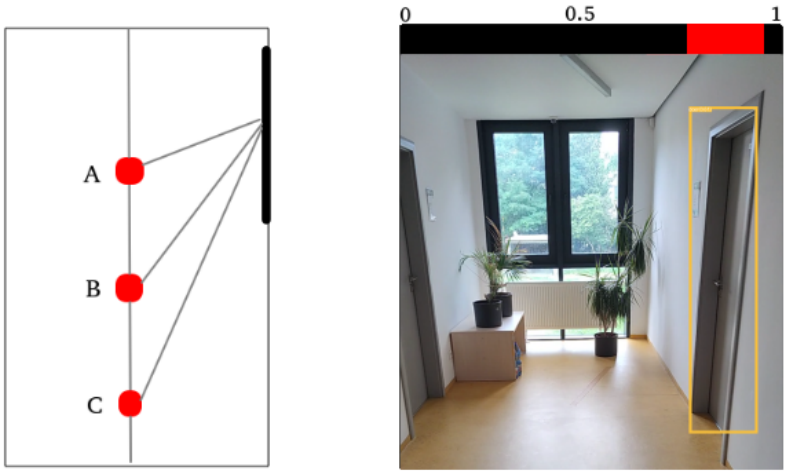

Figure 2: A door appears under different angles from positions A, B, and C (left image). The bounding box of the detected door determines the observed angle. This angle range is transformed into values between 0 and 1 (in this case, covering a 56-degree field of view, as indicated by the red rectangle). The proposed method matches this observed angle with the expected angle based on the user position.

- Doors are detected in the camera frame at the time of a detected step using Grounding DINO with the prompt *door*.
- Camera images may be modified using an anonymization technique, where humans are detected via Haar features and the corresponding regions are blurred.
- The grid-based filter performs a transition phase based on the detected step length and orientation, followed by a correction phase using the map model and door detection results. The position with the highest belief value is considered the estimated user position.

The grid-based filter estimates the user position based on the detected step. Each grid cell is assigned a belief value corresponding to the probability of the user's presence at that location. The sum of all belief values is equal to one, which is ensured after every step by normalizing the grid. In the evaluation described in the next chapter, a one-dimensional grid was used to represent a straight path. This simplification helps to highlight the impact of door detection. The proposed system is conceptually similar to the visualization by Fox et al. [25], but it employs discrete sampling of a continuous distribution.

The transition is performed based on the step length, as the orientation is considered constant in the experiment. The probability distribution is computed using a convolution with a mask derived from a normal distribution, where the mean corresponds to the estimated step length and the standard deviation is 10 cm.

The correction relies on two sources of information. The first is the map model, which prevents the user from moving through or beyond walls by setting the belief values of inaccessible locations to zero. For evaluation purposes, this correction is optional. However, in more complex scenarios, it may serve as the primary correction source. The second and main correction source is door detection. After processing, Grounding DINO provides a list of bounding boxes which cover the detected doors. Each detection includes the center position (x, y) of the object in the image, its size (width and height), and a confidence score. All values are within the interval $\langle 0, 1 \rangle$.

Figure 2 illustrates the door-matching process. For all grid positions, angles toward all doors are calculated. In a controlled scenario, only visible doors need to be considered, which requires knowledge of the smartphone camera's field of view. In the one-dimensional case, the angle is measured between the path and the door. For two-dimensional grids, angles are computed relative to the current device heading. Each angle is then mapped to the $\langle 0, 1 \rangle$ interval (with values possibly outside this range for objects outside the field of view). This mapped value serves as the mean of a normal distribution with a standard deviation of 0.2. A detected door is compared with all reference doors, and the probability of a

match is derived from this distribution.

More formally, the belief value for a grid cell x at iteration t is computed as follows:

$$B(x)^t = \sum_{x' \in G} B(x')^{t-1} M_l(x, x') \prod_{d \in D_t} \prod_{e \in E} P(d, e) C(d), \forall x \in G$$

where G is the grid containing all points (grid cells) on which the belief B is computed at iteration t. The function $M_l(x, x')$ returns the probability of transition from point x' to point x, with the index l denoting the mean of the normal distribution used for this calculation. D_t is the set of doors detected in the image at iteration t, and E represents the door model. In the implementation, only doors within the camera's field of view are considered. The probability of a match between a detected door d and a ground truth door e is computed using the function $P(d, e)$. The function $C(d)$ denotes the confidence score of the detected door d.

4. Evaluation

The goal of the evaluation is to demonstrate the capability of the door detection method to reduce uncertainty and correct errors accumulated by an imprecise PDR model.

4.1. Scenario

A straight corridor segment, 30 meters in length, is used for the evaluation. The first segment (12.35 meters) is traversed without applying door detection, allowing error accumulation. In the remaining segment, eight doors are visible. All doors are 90 cm wide but are unevenly distributed. On the right side, three doors are located at 16.0 m, 23.4 m, and 29.6 m from the starting point. On the left side, five doors are positioned at 16.1 m, 21.5 m, 24.4 m, 27.5 m, and 30.0 m. Measurements refer to the farthest point of each door. Additionally, there are 275 cm between the path endpoint and the wall enabling correction by the map model. The goal is to compute the difference between the true final position and the estimated position after completing the path.

The experiment was conducted twice: once in an empty corridor, and once with two people present (one standing, one moving). Initial experiments showed that the PDR system is reliable under these conditions: all 37 steps were detected, and heading deviations remained within acceptable limits. Therefore, a one-dimensional grid along the path is used for position estimation, and orientation is omitted. The grid resolution is 10 cm between adjacent cell centers. The dataset used in this evaluation is publicly available [26].

A Xiaomi MI 10 running Android 13 was used. A custom application recorded rear-camera video vertically, capturing a 56-degree field of view. The application also recorded timestamps for each video frame along with accelerometer and rotation vector measurements. Manual checkpoints were marked by the user and are included in the dataset.

Anonymization was applied using Haar feature detection for upper bodies, followed by blurring with a 33×33 kernel. In the first run (with no people present), the first right-side door was open. Duplicate door detections (i.e., overlapping bounding boxes) were removed, with the instance having the higher confidence retained for position estimation.

4.2. Door Detection

The exact evaluation of door detection performance is not the primary focus of this paper. The focus is placed on observing and understanding the behavior of the detection method, as opposed to quantitatively assessing its accuracy.

Not all eight doors were detected in any single camera frame. The best result achieved was the detection of three doors within one frame. The maximum distance at which a door was successfully detected from the camera was approximately 10 meters along the straight path. Figures 3 illustrate examples of detection outcomes.

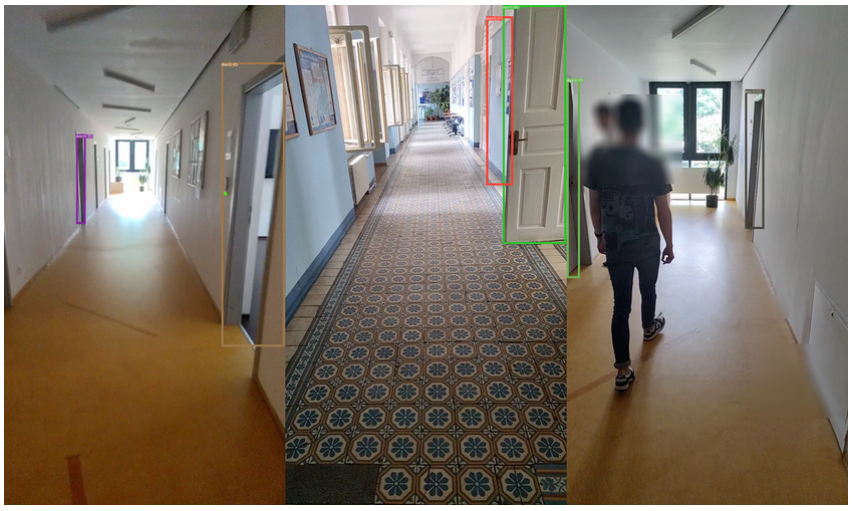

Figure 3: Door detection results are illustrated as follows: the first image shows a detected open door; the second presents a door opened into the corridor; and the third demonstrates detection in an image containing blurred individuals. First image is from Scenario 1 and third image is from Scenario 2 with anonymization.

Figure 4: Door detection examples: the first image shows duplicate detections of a single door; the second illustrates various objects incorrectly identified as doors; the third shows vending machines mistakenly detected as doors. Only first image is from Scenario 1.

The confidence scores of detected doors range between 30% and 69%. The lower bound is set by the threshold in Grounding DINO, for which the default value was used. Reducing this threshold may lead to the detection of additional doors.

The accuracy of door detection depends heavily on the characteristics of the building. In cluttered environments or in the presence of people, the detection performance decreases (see Figure 4). A frontal view of the door increases the likelihood of successful detection. Therefore, using the camera in horizontal orientation, which provides a wider field of view, may be advantageous. Some informal experiments also included open doors. When doors open into the corridor, they are more easily recognized.

In more cluttered buildings, there are more false detections, such as vending machines, whiteboards, and similar objects. However, false detection of windows occurred only under specific conditions. Overall, the results are satisfactory, with potential for further improvement through model fine-tuning, leveraging video sequences via multi-object tracking, or aggregating detection results across multiple frames rather than relying on a single image.

4.3. Positioning results

The true path was 30 meters long. PDR without a grid-based filter but with step length estimation resulted in an estimated distance of 33.68 m. Using a grid with a resolution of 10 cm, the estimated distance was 33.75 m. When the map model correction was applied, the estimated distance improved to 33.35 m. Thus, the estimation error was 3.75 m without the map model and 3.35 m with it. A small portion of this error is due to space discretization; however, the primary source of inaccuracy is the incorrect estimation of step length.

The PDR model tends to overestimate the actual step length. By incorporating the door detection method, the model was able to mitigate this inaccuracy. Particularly in the middle of the door segment, where three doors were successfully identified, the position remained nearly constant for three consecutive steps. This can be interpreted as the PDR model projecting a faster pace, while the door detection constrained the movement to align with the map and visual input from the camera. A summary of results is presented in the following table.

Errors in meters	Scenario 1	Scenario 2	Scenario 2 + anonymization
Grid	3.75	5.25	5.25
Grid + Map correction	3.35	3.35	3.35
Grid + Door detection	1.15	3.45	3.45
Grid + Door detection + Map	1.15	3.25	3.35

Scenario 1 was conducted in an empty corridor, while Scenario 2 included the presence of two people. Door detection refers to the proposed system. Map correction means that positions beyond walls are excluded from the grid. In this case, the distance between the endpoint and the last grid cell is 2.75 m; otherwise, it is 6.35 m.

These results are obtained in a specific scenario designed to exaggerate the impact of door detection, especially in Scenario 1. In a more complex scenario involving changes in direction, the map model would affect the accuracy more significantly. In this case, only a small correction is applied. However, door detection still enables position correction.

In Scenario 2, the characteristics of this path are demonstrated. If the step length is overestimated, the map correction improves the position. If it was underestimated, the map would have little effect for this path. However, the data suggests that door detection produces the same correction as the map model. Moreover, this applies even with anonymized images.

Conclusion

This paper presents an indoor positioning system based on Pedestrian Dead Reckoning (PDR) and a grid-based filter, where corrections are applied using door detection from camera images. The off-the-shelf model Grounding DINO is employed to obtain bounding boxes representing detected doors.

The evaluation demonstrates promising results, with the positioning error reduced from 3.75 m to 1.15 m using the proposed method. Additionally, an anonymization technique was incorporated to address privacy concerns related to the use of camera images.

These experiments suggest the possibility of using doors as visual landmarks for PDR correction. In the tested scenario, it exceeded, or at least matched, the corrections provided by the floor plan model. In such a small-scale experiment, the map model may seem obsolete. However, further experiments in more complex scenarios are necessary to determine whether this method provides similar corrections. It is also necessary to evaluate whether it can be considered an alternative to the map model or a complementary approach.

The main drawback of this method is its computational cost. The usage of text-to-image model Grounding DINO demonstrates the feasibility of using deep learning for door detection. Depending on the context, a fine tuned robust model, or a solution tailored to a specific building with an emphasis on efficiency and accuracy, would be preferable.

Moreover, several questions remain open. In particular, the feasibility of deploying such a system in real-time scenarios is yet to be confirmed. Furthermore, the effectiveness of the proposed correction method should be compared to corrections based on Wi-Fi or Bluetooth signals. Further work is also needed to improve the robustness and accuracy of the door detection process itself.

Acknowledgments

This research was supported by the Slovak Recovery and Resilience Plan, funded by the European Union – NextGenerationEU, under the project "Competence Center for Cybersecurity at Pavol Jozef Šafárik University in Košice", project code: 17R05-04-V01-00007.

Declaration on Generative AI

During the preparation of this work, the author used GPT-4 in order to: Grammar and spelling check and paraphrase and reword. After using these services, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking, IEEE Transactions on signal processing 50 (2002) 174–188.

[2] M. Opiela, F. Galčík, Grid-based bayesian filtering methods for pedestrian dead reckoning indoor positioning using smartphones, Sensors 20 (2020) 5343.

[3] T. Fetzer, F. Ebner, M. Bullmann, F. Deinzer, M. Grzegorzek, Smartphone-based indoor localization within a 13th century historic building, Sensors 18 (2018) 4095.

[4] V. Renaudin, M. Ortiz, J. Perul, J. Torres-Sospedra, A. R. Jiménez, A. Perez-Navarro, G. M. Mendoza-Silva, F. Seco, Y. Landau, R. Marbel, et al., Evaluating indoor positioning systems in a shopping mall: The lessons learned from the ipin 2018 competition, IEEE Access 7 (2019) 148594–148628.

[5] F. Potorti, S. Park, A. Crivello, F. Palumbo, M. Girolami, P. Barsocchi, S. Lee, J. Torres-Sospedra, A. R. J. Ruiz, A. Perez-Navarro, et al., The ipin 2019 indoor localisation competition—description and results, IEEE access 8 (2020) 206674–206718.

[6] F. Potortì, A. Crivello, S. Lee, B. Vladimirov, S. Park, Y. Chen, L. Wang, R. Chen, F. Zhao, Y. Zhuge, et al., Offsite evaluation of localization systems: Criteria, systems, and results from ipin 2021 and 2022 competitions, IEEE Journal of Indoor and Seamless Positioning and Navigation 2 (2024) 92–129.

[7] G. M. Mendoza-Silva, J. Torres-Sospedra, J. Huerta, A meta-review of indoor positioning systems, Sensors 19 (2019) 4507.

[8] G. Pipelidis, N. Tsiamitros, C. Gentner, D. B. Ahmed, C. Prehofer, A novel lightweight particle filter for indoor localization, in: 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2019, pp. 1–8.

[9] F. Galčík, M. Opiela, Grid-based indoor localization using smartphones, in: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2016, pp. 1–8.

[10] I. Klein, Pedestrian inertial navigation: An overview of model and data-driven approaches, Results in Engineering (2025) 104077.

[11] M. Vezočnik, M. B. Juric, Average step length estimation models' evaluation using inertial sensors: A review, IEEE sensors journal 19 (2018) 396–403.

[12] X. Yang, Y. Tian, Robust door detection in unfamiliar environments by combining edge and corner features, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE, 2010, pp. 57–64.

[13] M. Talebi, A. Vafaei, A. Monadjemi, Vision-based entrance detection in outdoor scenes, Multimedia Tools and Applications 77 (2018) 26219–26238.
[14] R. Kaur, S. Singh, A comprehensive review of object detection with deep learning, Digital Signal Processing 132 (2023) 103812.
[15] M. Antonazzi, M. Luperto, N. A. Borghese, N. Basilico, Development and adaptation of robotic vision in the real-world: the challenge of door detection, arXiv preprint arXiv:2401.17996 (2024).
[16] T. Zhang, J. Li, Y. Jiang, M. Zeng, M. Pang, Position detection of doors and windows based on dspp-yolo, Applied Sciences 12 (2022) 10770.
[17] G. Jocher, et al., ultralytics/ultralytics: Ultralytics yolo docs, https://github.com/ultralytics/ultralytics, 2023. Accessed: 2025-07-14.
[18] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: European conference on computer vision, Springer, 2014, pp. 740–755.
[19] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu, et al., Grounding dino: Marrying dino with grounded pre-training for open-set object detection, arXiv preprint arXiv:2303.05499 (2023).
[20] P. Viola, M. J. Jones, Robust real-time face detection, International journal of computer vision 57 (2004) 137–154.
[21] N. Ruchaud, J.-L. Dugelay, Automatic face anonymization in visual data: Are we really well protected?, in: Electronic Imaging, 2016.
[22] H. Hukkelås, F. Lindseth, Deepprivacy2: Towards realistic full-body anonymization, in: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, pp. 1329–1338.
[23] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, V. Lempitsky, Resolution-robust large mask inpainting with fourier convolutions, arXiv preprint arXiv:2109.07161 (2021).
[24] C. Villien, A. Frassati, B. Flament, Evaluation of an indoor localization engine, in: 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), IEEE, 2019, pp. 1–8.
[25] V. Fox, J. Hightower, L. Liao, D. Schulz, G. Borriello, Bayesian filtering for location estimation, IEEE pervasive computing 2 (2003) 24–33.
[26] M. Opiela, Supplementary evaluation files for the paper: Doors as visual landmarks for indoor positioning, Zenodo, 2025. doi:10.5281/zenodo.16354883.